Murray E. Taylor

McGraw-Hill
ELECTRICAL AND ELECTRONIC
ENGINEERING SERIES

FREDERICK EMMONS TERMAN, *Consulting Editor*

ELECTRONICS IN ENGINEERING

McGraw-Hill Electrical and Electronic Engineering Series

FREDERICK EMMONS TERMAN, *Consulting Editor*

BAILEY AND GAULT · Alternating-current Machinery
BERANEK · Acoustics
BRUNS AND SAUNDERS · Analysis of Feedback Control Systems
CAGE · Theory and Application of Industrial Electronics
CUCCIA · Harmonics, Sidebands, and Transients on Communication Engineering
EASTMAN · Fundamentals of Vacuum Tubes
EVANS · Control-system Dynamics
FITZGERALD AND KINGSLEY · Electric Machinery
GEPPERT · Basic Electron Tubes
GLASFORD · Fundamentals of Television Engineering
HAPPELL AND HESSELBERTH · Engineering Electronics
HARMAN · Fundamentals of Electronic Motion
HESSLER AND CAREY · Fundamentals of Electrical Engineering
HILL · Electronics in Engineering
JOHNSON · Transmission Lines and Networks
KRAUS · Antennas
KRAUS · Electromagnetics
LEPAGE · Analysis of Alternating-current Circuits
LEPAGE AND SEELY · General Network Analysis
MILLMAN AND SEELY · Electronics
ROGERS · Introduction to Electric Fields
RÜDENBERG · Transient Performance of Electric Power Systems
SEELY · Electronic Engineering
SEELY · Electron-tube Circuits
SEELY · Radio Electronics
SISKIND · Direct-current Machinery
SKILLING · Electric Transmission Lines
SKILLING · Transient Electric Currents
SPANGENBURG · Vacuum Tubes
STEVENSON · Elements of Power System Analysis
TERMAN · Electronic and Radio Engineering
TERMAN AND PETTIT · Electronic Measurements
THALER · Elements of Servomechanism Theory
THALER AND BROWN · Servomechanism Analysis
THOMPSON · Alternating-current and Transient Circuit Analysis
TRUXAL · Automatic Feedback Control System Synthesis

ELECTRONICS IN ENGINEERING

by W. RYLAND HILL, M.S., E.E.

ASSOCIATE PROFESSOR OF ELECTRICAL ENGINEERING
DEPARTMENT OF ELECTRICAL ENGINEERING
UNIVERSITY OF WASHINGTON

New York Toronto London
McGRAW-HILL BOOK COMPANY, INC.
1949

ELECTRONICS IN ENGINEERING

PREFACE

What? Another book on electronics?

Yes, this is a book on electronics, but it is one designed to fill a definite need—that of the advanced engineering or science student who wishes to learn something of electronics and its applications to his problems. Invariably such a student can devote only a limited amount of time to studies outside his special field, yet he wishes to discover the capabilities of electronic techniques and he needs to have a general understanding of how electronic instruments work in order to use them effectively. He needs to realize that the cathode-ray oscilloscope supplants older mechanical indicators, that amplifiers and vibration pickups have their limitations, and that rectifiers can provide large amounts of direct current more effectively than rotating equipment. And, fortunately, he can learn these things in a limited time by building upon the already relatively high technical level reached by him in his major field.

The material in this book results from the seven-year development of an elective course in electronics for nonelectrical engineering students at the University of Washington. The choice of subjects comes from questions and discussions with such classes and from having worked with many of the students on their individual problems of instrumentation and control in the fields of chemical, aeronautical, civil, and mechanical engineering, and also in physiology. Likewise the type of treatment results from an observation of the abilities of such students to understand the subject material.

The book contains sufficient material for a one-quarter or one-semester course on electronics, and it is divided into a number of relatively short chapters to facilitate selection to suit the needs of any particular group. For example, the chapter on polyphase rectifiers might be of particular benefit to chemical and mechanical engineers, while the chapter on strain gages and transducers is of particular interest to civil and aeronautical engineers. Perhaps

a group not interested in gas tubes and control circuits could omit Chaps. 5, 6, 7, and 8 without disturbing the continuity of the remainder. Flexibility such as this is a necessity in a course that presents selected material to a group of men studying outside their major fields.

The arrangement attempts to follow the introduction of each new electronic device by an application to illustrate its possibilities. Rather than studying diodes, triodes, pentodes, gas-filled tubes, etc., before discussing any applications, the chapter on single-phase rectifiers immediately follows diodes, the triode chapter contains the elements of amplification, and the chapters on electronic control circuits and polyphase rectifiers immediately follow the discussion of gas-filled tubes and phototubes. However, the individual instructor can easily alter this arrangement by assigning the chapters in a different order.

Comments and criticisms of the text are earnestly solicited. Does the graphic approach get across, or would a more mathematical one appeal to the students? Is the treatment too elementary, or is it more detailed than desirable? What additional subject matter should be incorporated? What might be left out?

Many thanks are due Michael Guidon of the University of Washington Mechanical Engineering Department for having struggled through much of the text in preliminary form to test its readability and to check many of the example problems.

W. Ryland Hill

Seattle, Wash.
October, 1949

CONTENTS

CHAPTER 1

ELECTRONS, IONS, AND PHOTONS

THE SCIENCE of electrical engineering deals with the movement of electrons and other charged particles, together with their associated electric and magnetic fields. For the first 60 years of its approximately 100-year life, electrical engineering dealt primarily with the flow of electrons in conductors. During this period were developed the telegraph and early telephone, the electric lamp, and the electric motor and generator. These were important devices, destined to speed communications, lengthen the day, and provide flexible power to supplement the already important mechanical-engineering advances.

By 1904 the electrical industry was well on its way to becoming an important factor in the economics of this country. In this year came an important step in a new direction—a first step in the field of electronics. Fleming, in England, patented a device known as the Fleming electric valve and used it for the detection of wireless signals. Two years later De Forest added a grid to control the flow of current through the valve. The basic importance of this step is that *for the first time the engineer could control the flow of free electrons in space.* This added to the purely power aspects of electrical energy the delicate control functions of the electric valve which permitted the development of radio, long-distance telephony, television, electronic control, and radar.

Before attempting to study the particular devices employing electrons, ions, and photons, we must first direct our attention to the properties of these physical entities.

1.1 Atomic Structure

The geometric concept of the atom is a familiar one to every student who has taken college physics. This concept pictures a

small dense cluster of protons and neutrons which carries all but a tiny fraction of the mass of the atom. Surrounding this nucleus a number of satellite electrons spin about in various circular and elliptical orbits. The number of electrons in this outer part of the atom equals the number of protons in the nucleus and makes zero the net electrical charge. These electrons are pictured as tiny light bits of negative electric charge.

Figure 1.1 shows a geometric picture of an atom of boron together with a tabulation of its particles. The chemical and electronic reactions of an atom are entirely dependent upon the behavior of the satellite electrons; hence it is these that are of primary interest to us.

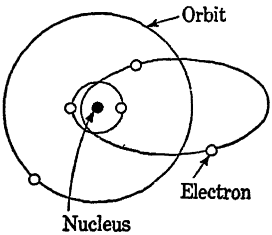


FIG. 1.1. The geometrical picture of an atom of boron.

Name of particle	Number of particles	Charge on each one	Approximate mass
Electron.	5	−1	$\frac{1}{1,850}$*
Proton ..	5	+1	1
Neutron.	6	0	1
Total .	16	0	11

*Mass in atomic units of 1/16 oxygen atom.

Now this satisfying picture is a convenient one for explaining many of the properties of electron tubes, but it must be pointed out that no one has ever seen an atom. The picture drawn above is of use only in providing a convenient physical framework for relating in a reasonable manner the various phenomena that have been observed by physicists. More complete investigations indicate its inadequacy. Electrons under many circumstances seem endowed with wave properties; their atomic orbits are always an integral number of wavelengths long, and this restricts the number of permissible orbits in an atom.

Energy Levels. Another convenient atomic picture is one showing the energy levels represented by the various possible electron orbits. Figure 1.2 illustrates such a diagram for a hydrogen atom. This atom contains but one nuclear proton and a single electron. The electron seeks the shortest possible orbit close to the nucleus and stays there unless disturbed. This is called the *normal* state of the atom represented by the base line of the diagram. The energy level corresponding to this orbit has been chosen as zero and given the *quantum number* 1. Somewhat farther out from the nucleus exists another orbit which fits the wavelength requirements of the electron. This orbit carries the quantum number 2. Since the nucleus attracts the electron, work

is required to remove it to a greater distance; consequently the atom possesses greater energy than before. When the electron has somehow been moved out to this position, the atom is said to be in the *excited state*, and the energy required to move it out

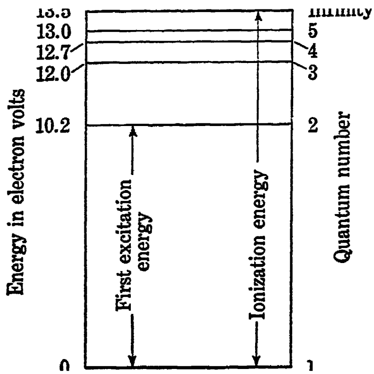


FIG. 1.2. Energy levels in a hydrogen atom.

to this next orbit is called the first excitation energy. No smaller amount of energy can be accepted by a normal atom.

Still farther out exist other permissible orbits with integral quantum numbers 3, 4, 5, etc., up to infinity, which corresponds to completely removing this electron from the atom. The diagram shows the corresponding energy levels. Ordinarily the atom does not remain in an excited state an appreciable amount of time; the electron immediately drops back to some lower level with the excess energy given off as radiant energy or light.

Ionization. The process of completely removing an electron from an atom is called ionization, and each particular type of atom requires a definite amount of work to do this called the ionization energy or *ionization potential.* The ionization of hydrogen requires 13.5 electron volts of energy. An electron volt is equivalent to the work required to move a single electron charge against a potential difference of 1 volt. It equals 1.6×10^{-19} joule. Complex atoms with large numbers of electrons can be singly, doubly, or triply ionized as one, two, or even three electrons are removed. In electron tubes, however, multiple ionization rarely occurs.

The energy required to excite or ionize an atom can come from a

collision with another electron, an ion, or an atom. Such collisions are classified as either *elastic* or *inelastic*. If no excitation or ionization takes place and the particles have the same total kinetic energy as before, the collision is called elastic. This is invariably the case when the colliding particles carry less energy than that required for the first excitation level. When the particles carry more than this minimum energy, the collision may be inelastic but the probability of excitation or ionization is not more than about 0.2 percent, and it may be much less. A single fast electron passing a short distance through a gas, however, makes many collisions with a good chance of producing one or more ions.

The ionization or excitation energy can also come from electromagnetic radiation or light.

Photons. All early experimental studies of light, such as diffraction and interference, were large-scale observations which clearly indicated the wave structure of electromagnetic radiation. More recent observations of the behavior of light in atomic reactions just as clearly show light to behave as particles consisting of elementary chunks of electromagnetic radiation called photons. Since under one set of circumstances light behaves as particles, while under another it seems to be waves, we are at liberty to use whichever concept is most convenient.

The experimental evidence indicates that each photon carries a definite amount of energy called a quantum. This energy is related to the light frequency by the expression

$$w = hf \qquad \text{joules} \tag{1.1}$$

where w = energy of quantum
h = Planck radiation constant (6.6×10^{-34} joule-second)
f = frequency of radiation (cycles per second)

When a photon with sufficient energy strikes an atom, it may produce excitation or even ionization with the release of an electron. After the collision the photon may proceed in a new direction with less energy than before which, if Eq. (1.1) is correct, must mean that the frequency has been reduced. This fact has been carefully verified for X rays. Under other circumstances the photon expends all its energy and vanishes, and the electron carries off the excess in kinetic form.

An excited or ionized atom usually does not remain long in this state but soon returns to the normal state of lower energy. If

ionized it attracts any convenient stray electron, a process called recombination. If excited, the electron almost instantly returns to its normal orbit. In either of these processes the excess energy is given off in the form of photons. As an example, let us study the return of a single hydrogen electron from the lowest state of excitation to the normal state. Reference to Fig. 1.2 shows an energy loss of 10.2 electron volts. The charge on an electron is 1.6×10^{-19} coulomb so that 10.2 electron volts represents an energy of

$$w = (10.2)(1.6 \times 10^{-19}) = 1.63 \times 10^{-18} \text{ joule}$$

Placing this value of w in Eq. (1.1) and solving for the frequency, we obtain

$$f = \frac{(1.63 \times 10^{-18})}{(6.6 \times 10^{-34})} = 2.47 \times 10^{15} \text{ cps}$$

This frequency corresponds to a line with a wavelength of 1,216 angstroms found in the ultraviolet hydrogen spectrum.

1.2 Electrons in Solids

To ionize or remove an electron from an isolated atom always requires a definite amount of energy, as illustrated in Fig. 1.2. In a gas the atoms are so well separated that the ionizing potential is not appreciably affected by the presence of the other gas atoms. In a solid, however, the atoms are so closely spaced that the electrons in the outer orbits are influenced by the electric fields of the adjacent atoms. In metals and a few other solids the resulting confusion of fields decreases the bond between the outermost electrons and the nucleus to the point where they can travel freely from one atom to another.

Conduction. Materials in which these free electrons occur are the electrical conductors; applying an electric potential difference between the ends of a conductor causes the free electrons to drift toward the positive end and produce an electric current. In most materials this flow is proportional to the impressed voltage as expressed in the familiar Ohm's law of electric circuits. The heat produced by current flow in a conductor results from the collisions or, more properly, interactions of the moving electrons with the atoms of the solid. The kinetic energy imparted to the atoms thus increases their energy of vibration and raises the temperature.

Nonconductors are those materials which do not contain ap-

preciable numbers of free electrons. However, if such materials are heated, the increased kinetic disturbance may give some of the outer electrons sufficient energy to escape their force bonds and travel through the material under the influence of an applied voltage. It is generally true that insulators become better conductors at high temperatures. The energy for freeing electrons in a poor conductor may also come from light. The conductivity of selenium, for instance, increases with illumination. This is explained by assuming that the light photons give some of their energy to produce free conduction electrons in the material. One type of photosensitive device operates on this principle.

1.3 Electron Emission

Under certain circumstances the free electrons may escape from the surface of a conductor. This is called *electron emission.* It is of importance because almost all electron tubes depend upon the emission of electrons for their operation.

Work Function. To remove an electron from a metal requires the expenditure of a definite amount of work. From a qualitative standpoint, the removal of an electron leaves behind a local positive charge which attracts the electron being removed. Consequently this attractive force resists the removal, and work must be done. This work measured in electron volts is called the work function of the material.

The work functions of common materials vary from about 6 volts for platinum to a low of 1.8 volts for cesium. Composite surfaces with an extremely thin layer of one metal on another may possess even lower values of work function. An example is the oxide-coated cathode used in vacuum tubes. This surface has a work function of about 1 volt.

Thermionic Emission. Kinetic theory pictures all the atoms of a solid in a constant state of agitation about an average position. At high temperatures the energy of agitation increases until the forces binding the atoms into position are overcome and the material first melts and finally evaporates. In conductors the free electrons also partake of this heat motion and continually drift around in random fashion. Some of these electrons have very low velocities, while others, because of a particularly fortuitous series of collisions, may reach high speeds. The mean velocity increases with the temperature.

Any electrons which possess sufficient velocity directed toward the surface may have enough energy to escape. At normal temperatures the number of electrons with a kinetic energy larger than the work function is practically zero, but at a temperature above 1000°K a useful amount of emission can be obtained. This type of emission is called *thermionic* emission because the electrons obtain their energy from heat. In practical electron tubes the heat energy is produced electrically by passing current through a resistance wire.

Field Emission. Field emission refers to the emission of electrons from a metal surface due to the attraction of a nearby positive electrode. This attraction offsets the normal affinity of the emitting surface for its own electrons, and emission from a perfectly cold metal surface can be obtained with high field intensities in the order of 10^8 volts per meter. In ordinary low-voltage tubes field emission does not occur, but high-voltage X-ray equipment must be carefully designed to prevent undesired field emission. The necessary precautions include large electrode spacings and careful rounding of all sharp edges.

Secondary Emission. Secondary emission is the emission of electrons that have gained kinetic energy from some primary bombarding particle, usually another electron. It occurs at any metal surface subjected to bombardment by either ions or electrons. A high-velocity primary particle striking the surface of a metal gives up some of its kinetic energy to one or more electrons in the immediate vicinity, and those receiving energy may escape. As might be expected, secondary emission takes place most readily from materials with low work functions. The process is surprisingly effective, and a single primary electron may produce as many as 10 secondaries from a specially prepared surface. Positive ions are much less effective than this, and the probability of producing a secondary electron is very much less than unity.

Although secondary emission is used to advantage in some types of vacuum tubes, it is often a disadvantage, and special precautions must be taken to avoid it.

Photoelectric Emission. Electron emission can also be obtained by bombarding a metal surface with photons of light. Incident photons give up their energy to electrons near the surface, and if sufficient energy is available, emission takes place. The possibility of emission can be expressed by the following equation set forth

by Einstein in 1905:

$$\frac{mv^2}{2} \leqq hf - W \tag{1.2}$$

where $mv^2/2$ = maximum kinetic energy of emitted electron, in joules (m in kg, v in m per sec)

hf = incident photon energy as computed from Eq. (1.1), in joules

W = work required for electron escape, in joules

This equation indicates that the maximum possible kinetic energy of an emitted electron is equal to the incident energy minus the energy required for escape. The majority of the electrons will have smaller energies than the maximum because of internal collisions or because the velocity is not directed toward the surface. However, Eq. (1.2) does predict that no emission can take place unless hf is larger than W. This requires high-frequency (ultraviolet) light for most common metals, but especially prepared surfaces with low work functions respond to the lower visible and infrared frequencies.

The amount of emission depends upon the ability of the photons to transfer energy to the electrons. This depends on the surface characteristics and on the frequency of the light. For a given light frequency, however, the current is accurately proportional to the light intensity. Commercial photosensitive surfaces provide emission currents of several microamperes per square centimeter under an illumination of 100 foot-candles.

PROBLEMS

1.1 Compute the frequency and wavelength of light given off when hydrogen ions recombine with electrons to produce normal hydrogen atoms. The velocity of light is 3×10^8 m per sec.

1.2 The shortest wavelength found in the mercury spectrum is 1,188 A. Compute the ionization potential of mercury.

1.3 Compute the minimum velocity that an electron must have to ionize mercury vapor. Electron mass is 9.1×10^{-31} kg.

1.4 An experiment to determine the work function of a metal shows that red light does not produce photoemission but that any blue light having a wavelength less than 4,500 A does. What is the surface work function expressed in volts?

1.5 Visible light containing all wavelengths between 4,000 and 8,000 A falls on a photoemissive surface having a work function of 1 volt. Compute the maximum possible kinetic energy possessed by any of the emitted electrons.

1.6 A sensitive type of phototube employs a 10-stage electron multiplier to increase the output current. In this multiplier the tiny photoemission current is directed toward a target at which each primary electron knocks out several secondary electrons. These in turn accelerate toward a second target to produce a further current increase, and so on. In a 10-stage multiplier this occurs 10 times, and the final output current is 200,000 times the input. Find the ratio of secondary emission to primary current at each bombardment.

CHAPTER 2

PRACTICAL EMITTERS AND DIODES

THE SIMPLEST electron tube is the *two-element diode* consisting of a hot *cathode* for emitting the electrons and a positive *anode,* or plate, for collecting the electron stream. This device has the interesting property of acting as a one-way valve permitting current to flow through it when the anode is positive. With reversed polarity, the electrons are repelled back to the cathode and no current flows.

2.1 Thermionic Emission Law

According to kinetic theory, all the particles in a solid continually vibrate about an average position. This vibration is small at low temperatures and becomes increasingly violent when the temperature is raised. In a conductor, the free electrons also partake of this heat motion and dart at random through the metal. These free electrons have a velocity distribution ranging from very small velocities (for an electron that has undergone a head-on collision and nearly stopped for a moment) to relatively high ones (for a particle that has experienced an adventitious series of collisions all urging it in the same direction). Only a tiny fraction of the electrons have these high and low extremes of velocity; the majority have velocities clustering about a value called the mean velocity for all electrons.

At ordinary temperatures a negligible number of electrons possess sufficient kinetic energy to exceed the work function of a metal, and no thermionic emission can be detected. At higher temperatures, however, the velocities of all particles increase, and the number of electrons that can escape becomes appreciable. This concept of thermionic emission is similar to the kinetic picture of the evaporation of liquids, and it is not surprising to find that the law deduced from a statistical study of free electrons in metals leads to a law of emission very like the law of evaporation. Sev-

eral such laws have been formulated, of which the most modern is Dushman's law

$$I = AT^2e^{-b_0/T} \qquad \text{amp per sq cm} \tag{2.1}$$

where T = absolute temperature, in degrees Kelvin
b_0 = temperature equivalent of the work function
= (work function) (a constant)
A = a universal constant of 120.5

This law expresses well the experimental shape of the emission curve although there is considerable discrepancy between experimental and theoretical values of A. However, Eq. (2.1) does show the emitted current to be very sensitive to the temperature and the work function because both appear in the exponent. For low values of work function a reasonable emission can be obtained at a much lower temperature than with a high work function. This is illustrated by the curves of Fig. 2.1 which show the emission

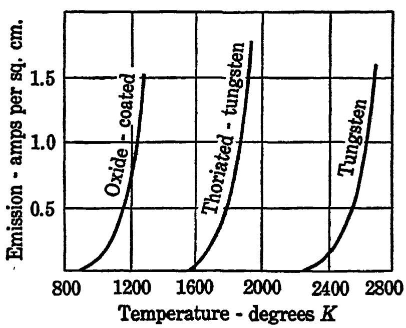


FIG. 2.1. Comparative emission characteristics of the three commonly used cathode materials.

characteristics of the three commercially practical thermionic emitting materials.

Even at the peak emission of a normal emitter the fraction of the available free electrons released is extremely small, and it is not necessary to consider depletion of the source.

2.2 Practical Emitters

A practical thermionic emitter should (1) provide a large emission current per watt of heating power, (2) possess mechanical strength at operating temperature, (3) be electrically stable during its useful life, and (4) possess a reasonably long life. Item 1 sug-

gests choosing a material with a low work function to permit easy electron escape at low temperatures and correspondingly small heating power. However, this property must be correlated with the ability to withstand advanced temperatures without softening. Unfortunately, the metals with low work functions are the alkaline earths (sodium, for example), which are soft and have low melting points.

Tungsten. Tungsten is the only pure metal that is practical for use as a thermionic emitter. It possesses every desirable quality listed above except the first; its work function is high (4.5 volts), and it must be operated at about 2500°K to obtain the best balance between maximum emission and reasonably long life. At this temperature the emission current is about 2 milliamperes per watt of heating power. During its life the tungsten filament slowly evaporates until, after several thousand hours, it fails either mechanically or from reduced emission.

Tungsten cathodes are employed in all high-voltage tubes where the operating conditions are so severe as to prohibit the use of more efficient but less stable materials.

Thoriated Tungsten. The search for a material with a lower work function but with the mechanical advantages of tungsten led to the development of the thoriated-tungsten filament. This is a composite structure consisting of a tungsten base with a thin, possibly monatomic, layer of thorium on the surface. This layer seems to evaporate from the tungsten less rapidly than from thorium itself, and the composite surface has an even lower work function than pure thorium. At a temperature of 2000°K, a thoriated surface provides an emission of about 100 milliamperes per watt of heating power, combined with a reasonable life expectancy.

Unfortunately, this much higher efficiency is obtained at the expense of stability under the effects of positive-ion bombardment. All vacuum tubes contain some residual gas which may be ionized by collisions with electrons speeding toward the anode. The positive ions accelerate toward the cathode and strike it with a kinetic energy proportional to the potential drop across the tube. This does no damage at low voltages, but above about 10 kilovolts the thorium surface is rapidly removed. Consequently, the inefficient but rugged tungsten must be used for high-voltage tubes.

A thoriated-tungsten filament is constructed by adding a small amount of thorium oxide to the tungsten powder before it is

sintered and drawn into wire form. After being mounted in the tube, the filament is heated to 2800°K to decompose the thorium oxide to thorium, after which the temperature is reduced to permit a small amount to diffuse to the surface and build up a thin layer. Modern thoriated cathodes are also carbonized by heating in a hydrocarbon vapor. It has been found that this stabilizes the surface and lengthens its life.

Oxide-coated Cathodes. The oxide-coated cathode is an even more efficient although less stable surface than thoriated tungsten. It is produced by coating a suitable metal surface with a mixture of barium and strontium carbonates. After mounting and evacuation, the cathode is heated to reduce the carbonates to oxides; probably a small amount of barium is liberated and distributed through the mixture. The resulting composite surface has an exceptionally low work function of about 1 volt and provides an emission efficiency above 100 milliamperes per watt at a temperature of only 1000°K.

The surface is sensitive to positive-ion bombardment and has the additional disadvantage that the spongy surface tends to release gas during the life of the tube. This increases the formation of positive ions and hastens the destruction of the cathode. Consequently, oxide-coated cathodes are not used in tubes operating at anode voltages much in excess of 600 volts. Since the majority of tubes used in radio receivers and laboratory electronic equipment operate at less than 600 volts, oxide-coated cathodes are the type most commonly encountered.

2.3 Cathode Construction

Tungsten and thoriated-tungsten cathodes operate at such high temperatures that the emitting surface must be *directly* heated.

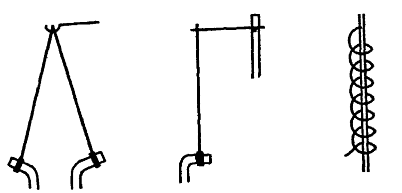


FIG. 2.2. Typical directly heated filaments.

This means that the cathode is in the form of a wire filament with the heating current passing through it. Figure 2.2 shows typical

types of filament construction. Figure 2.3 illustrates the conventional diagram for a diode with a directly heated cathode. The circle represents the envelope, the inverted *V* is the filament, and the upper bar is the anode for collecting the electrons. The filament is usually heated from a low-voltage a-c source.

The directly heated cathode has the disadvantage that the cathode and the heating source are electrically coupled; in sensitive amplifiers this may introduce hum. Fortunately, efficient oxide-coated cathodes operate at such a low temperature that it is

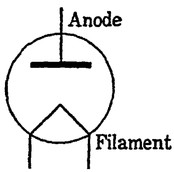


FIG. 2.3. Conventional diagram for a diode with directly heated cathode.

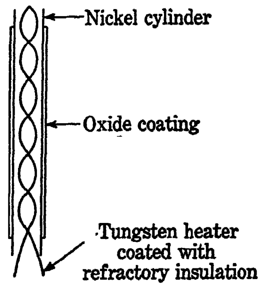


FIG. 2.4. Typical construction of an indirectly heated cathode.

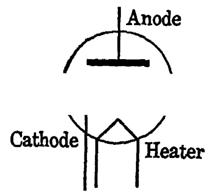


FIG. 2.5. Conventional symbol for a diode with an indirectly heated cathode.

possible to separate the functions of heater and emitting surface by employing an *indirectly heated* cathode. In a typical form, the oxides coat a small nickel cylinder heated by a hot tungsten wire placed inside the cylinder but insulated from it, as shown by Fig. 2.4. With this arrangement, several cathodes at different potentials can be heated from a single a-c source with the heaters connected in series or parallel. Most a-c radio receiving tubes employ indirectly heated cathodes. Figure 2.5 shows the conventional diagram for a diode with an indirectly heated cathode.

2.4 Vacuum Diode Characteristics

The electrical characteristics of a diode can best be understood by performing an imaginary experiment using the circuit of Fig. 2.6. We shall hold the filament voltage (and thus temperature) constant and vary the applied potential e_b to observe the effect on the anode current i_b. Observe that the arrow shows the *con-*

ventional direction of plate current flow which is opposite the actual direction of electron flow. This results from the definition of current, as the direction *positive* charges must move to produce the observed effects. Since an electron moving to the right has

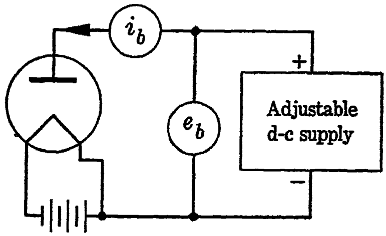


FIG. 2.6. Experimental circuit for measuring diode characteristics.

the same net electrical effect as an equal positive charge moving to the left, the conventional current and the electron flow are opposite. Because electrons carry the majority of the charge in electrical circuits, it is unfortunate that the early choice of positive and negative turned out as it did. Had the terms been reversed, many physical explanations would have been more convincing, although the mathematics of the process would not be appreciably simplified.

Figure 2.7 shows the resulting curves of plate current versus

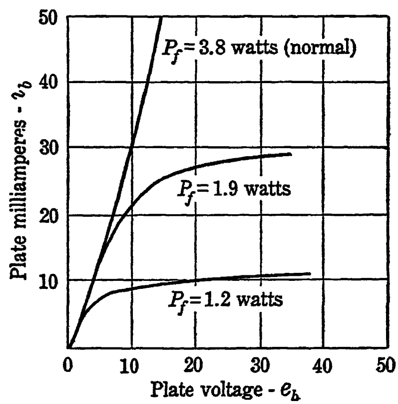


FIG. 2.7. Diode-characteristic curves for different values of cathode heating power. The flatter portion of each curve represents the condition of saturation.

plate voltage for several values of heating power. Each curve first rises at an increasing rate and then more or less levels off

at a value determined by the heating power. This suggests that the plate current is independent of the attractive power of the anode above a certain point. From this we reason that a limit is reached because the anode draws all the available electrons to itself, and further increases of potential can attract no more. Under this condition the anode current represents the actual cathode emission, and the tube is said to be temperature *saturated.* At low values of heating power there is less emission, and saturation takes place at a lower current. The condition of saturation has little practical interest for us because tubes are designed with a normal emission several times greater than any current the tube is expected to handle.

In the unsaturated region the curves are nearly identical, the anode current is *less* than the cathode emission, and some of the electrons return to the cathode despite the attraction of the positive plate. This is attributed to the negative space charge or electron cloud existing between the two electrodes. To help clarify this concept, Fig. 2.8 shows curves of the potential-space relation

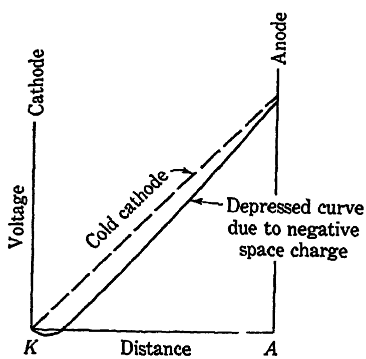


FIG. 2.8. Curves showing the potential distribution in a diode with plane parallel electrodes. The dotted line shows the distribution when the cathode is cold and emits no electrons. The heavy curve shows the potential depression due to space charge with a hot cathode and plate-current flow.

for an idealized diode with plane parallel electrodes. With a cold cathode and no electron flow, the electric field between the plates is uniform and the potential curve is a straight line. A single electron leaving the cathode "coasts" uphill toward the more positive region and can reach the anode. With a hot cathode, however, the heavy electron flow to the plate results in a large density of negative charges in the intermediate space. This de-

presses the potential, as shown by the solid curve. Now an electron leaving the cathode first experiences a downhill slope which hinders its flight; only those with sufficient initial kinetic energy can underride the trough and cross the gap. Since electrons leave the cathode with a range of energies (in the order of a few tenths of an electron volt), the fraction that can reach the plate depends on the plate potential and the cathode temperature. Surprisingly enough, the cathode temperature has relatively little effect because an increased emission produces a denser space charge and a more negative trough. A smaller fraction of the total emission can now underride the trough, and the net current increases only slightly.

From an engineering point of view, we are interested only in the top curve for normal heater power (Fig. 2.7) which shows no saturation within the rated current limitations of the tube. For negative anode voltages the plate current is zero; the anode repels all electrons back to the cathode. We thus have a device corresponding to a check valve that is an open circuit in one direction and a fairly good conductor in the other. Such a device is called a rectifier and can be used in any number of interesting applications, a few of which are discussed in Chap. 3.

2.5 Diode Construction

Figure 2.9 shows the construction of a typical diode. Surrounding the indirectly heated cylindrical cathode is an anode formed

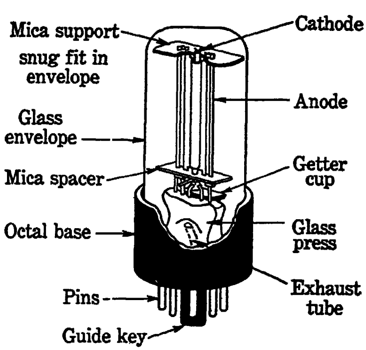


FIG. 2.9. Cutaway view of a type 35Z5 vacuum diode.

of two flat metal plates with semicylindrical grooves pressed in them. When spot-welded together, these plates form a cylinder with large fins to radiate the heat generated by electrons striking

the anode. Mica end supports accurately maintain the concentric anode-cathode spacing and hold the whole assembly firmly centered in the glass envelope.

At the bottom is the press formed of a piece of heated glass tubing squeezed together at the top end to seal around the incoming leads from the base. These wires are of an alloy having a coefficient of expansion matching that of the glass. Above the press they are spot-welded to the various leads and supports of the tube assembly. The lower part of the press is flared out to fit the tubular glass envelope. In making the tube the press is first constructed and then welded to the anode-cathode assembly. This is followed by slipping the envelope over the assembly and sealing the envelope and press together at the bottom. Next the air is exhausted through a glass tube connected to the press. While the pumps maintain a high vacuum, the interior parts of the tube are induction heated to drive off any occluded gases that might later be liberated to spoil the vacuum. The envelope is heated nearly to softening for the same purpose. After sufficient time the exhaust tube is heated to melting at one spot, and it collapses to seal off the tube.

As a last precaution the getter cup, which contains barium or magnesium, is heated to vaporize the metal and deposit it on the glass wall as a silvery coating. This acts as a scavenger to combine with any residual oxygen or nitrogen freed during the life of the tube. These precautions ensure a vacuum of about 10^{-8} atmosphere, and the probability of an electron striking a gas atom on its flight to the anode is extremely small.

The tube is completed by cementing on the proper base and connecting the lead wires to the pins. The figure shows a standard octal base which provides eight pins and a guide key for orienting the tube in the socket. In this particular case only four of the pins are active, the others being dummies. Standards for tube numbering, bases, pin connections, etc., have been adopted from time to time, but the rapid advance of the industry and the tremendous increase in number of tube types have caused each system to be outgrown a few years after it was formulated.

2.6 Contact Rectifiers

Another device having electrical characteristics similar to a diode is the contact rectifier, or barrier-layer cell. A barrier-layer cell

consists of a good conductor separated from a poor conductor by a barrier layer. The separation is more electrical than physical since the two materials make intimate contact at the boundary. One commercial version of this construction is the copper-oxide rectifier shown in cross section by Fig. 2.10. The good conductor is the copper disk on which is formed a layer of copper oxide in intimate contact with the mother copper. A lead washer makes contact with the back of the oxide layer, with the electrical con-

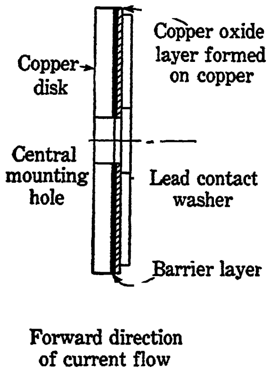


FIG. 2.10. Sectional view of a copper-oxide contact rectifier.

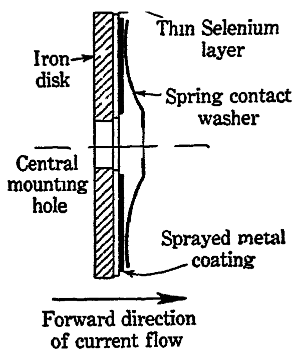


FIG. 2.11. Sectional view of a selenium contact rectifier.

nections brought out from the copper and the washer. A more recent and more efficient type of contact rectifier is the selenium cell consisting of an iron disk coated with a layer of selenium, as shown by Fig. 2.11. The selenium face is sprayed with a thin layer of low-melting-point solder and then given an electrical forming treatment to produce a barrier layer between the coating and the selenium. The iron serves as a support and an electrical contact to the selenium corresponding to the lead washer in the copper-oxide rectifier.

Both of these devices exhibit a very high resistance to the flow of current in one direction and a low resistance to flow in the other direction (called the forward direction). The theory of these cells is rather complex, but it is supposed that the peculiar conduction properties are due to the abundance of electrons in the good conductor as compared with the relative scarcity of free electrons in the poor conductor on the other side of the barrier layer.

Figure 2.12 shows the properties of a typical selenium rectifier having an effective surface area of 1 square centimeter. To obtain the curve for a disk with a larger area, merely multiply the ordinates of Fig. 2.12 by the area in square centimeters. For instance, a disk with an area of 50 square centimeters can carry a forward current of about 2 amperes with a voltage drop of only 1 volt. In the reverse direction a single cell can stand a voltage of about 25 volts maximum before reverse current becomes serious. In

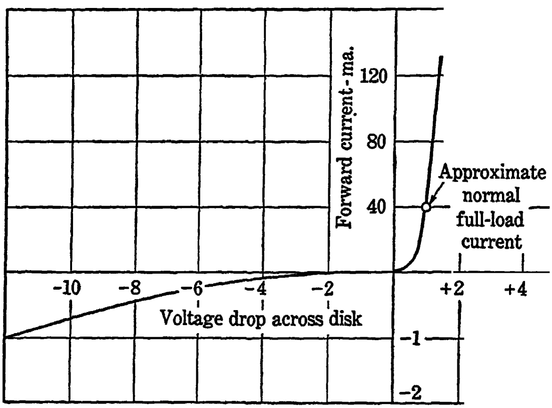


FIG. 2.12. Average characteristics for a selenium contact rectifier having an effective area of 1 sq cm.

this it differs from the diode because the reverse current in a diode is zero.

From the standpoint of circuit design, the contact rectifier acts much like a diode except that it passes a relatively large current with a very small voltage drop. Vacuum diodes, on the other hand, are inherently low-current devices with considerably larger voltage drops. Fortunately, vacuum diodes can withstand higher reverse voltages so that they are used in high-voltage circuits where the voltage drop when conducting is not a serious fraction of the applied voltage. Low-voltage circuits (up to 100 volts or so) designed for currents over a few tenths of an ampere would probably employ contact rectifiers. The contact rectifier also has the advantage of requiring no inefficient hot cathode, and it is instantly ready for service without requiring time to heat up.

A number of disks can be operated in parallel for increased cur-

rent capacity up to several hundred amperes. They can also be connected in series to withstand reverse voltages higher than 25 volts.

PROBLEMS

2.1 The current flow in the electrolyte of a battery consists of positive ions flowing to the right and negative ions flowing to the left. The flow rate is 10^{20} ions per second for each type, and each ion carries a single electronic charge. Find (*a*) the net current flow, and (*b*) the direction of conventional current flow.

2.2 In a diode with plane parallel electrodes the electrons flow in parallel paths to the anode. The diode current is 100 ma per sq cm of electrode area. At a point near the cathode the electrons have a low velocity of 10^6 m per sec, but when they strike the anode they are traveling 10 times as fast. Compute the electron density in number of charges per unit volume at each of the two positions.

2.3 An electron leaves the cathode of a diode with practically zero velocity and travels to the anode 100 volts more positive. Compute the velocity with which it strikes the anode. (*Note:* Work $= VQ$ joules where V is the potential difference in volts through which charge Q coulombs moves.)

2.4 The diode of Prob. 2.3 has plane parallel electrodes spaced 0.2 cm apart so that the electron experiences a uniform acceleration throughout its flight. Compute the transit time taken for the electron to pass from cathode to anode.

CHAPTER 3

SINGLE-PHASE RECTIFIER CIRCUITS

AN IMPORTANT application of diodes and contact rectifiers is in the production of direct current from an a-c power source. Rectifiers are more convenient than batteries for small amounts of power, and for large power requirements they may be even more efficient than motor-generator sets.

Rectifiers providing small currents at 50 volts or above commonly employ simple vacuum diodes because efficiency is not particularly important. Circuits for moderate voltages and large amounts of power employ gas-filled tubes because of their low voltage drop and high efficiency. Such tubes are discussed in Chap. 5. Low-voltage circuits providing more than a few tenths of an ampere are usually designed for contact rectifiers.

3.1 Half-wave Rectifier with Smoothing Capacitor

One simple basic rectifier circuit is the half-wave rectifier shown in Fig. 3.1. This arrangement is used for small amounts of power when simplicity is the most important consideration. In this circuit, transformer T isolates the circuit from the a-c line and provides the desired magnitude of alternating voltage for rectification. Capacitor C serves to smooth the output voltage, and R_L represents the load on the circuit. The actual load often consists of an electrical circuit drawing an equivalent amount of current. This rectifier circuit usually employs a vacuum diode as shown, but a contact rectifier could also be used. The direction of the arrow represents the only direction in which current can flow through the element.

The operation of this circuit is a typical example of one in which discontinuous action takes place because of the presence of a unilateral, or "one-way," device. Figure 3.1 shows the input to the diode represented as a sine-wave voltage e_1. Let us imagine that the circuit has been energized at the moment shown by

point a. We shall further assume that in any practical and efficient circuit the voltage drop in the diode is small compared with the output voltage. Thus during the first positive quarter cycle after

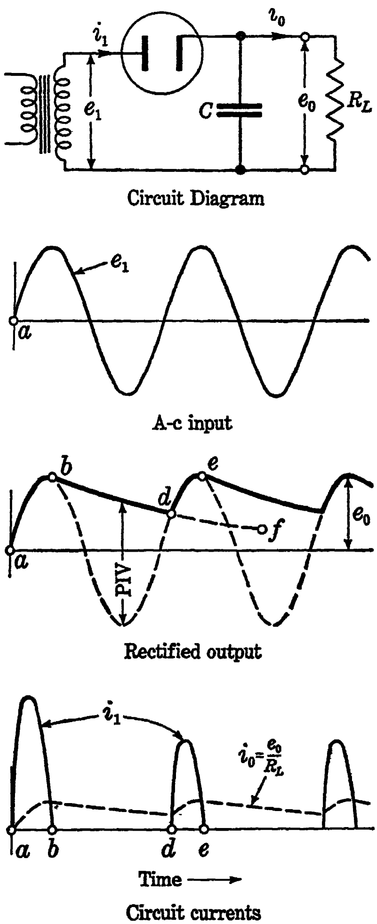


FIG. 3.1. Circuit diagram and wave forms for a half-wave rectifier with smoothing capacitor.

time a the output voltage e_0 rises along with the input until point b is reached. This can happen because the diode conducts well in the direction of the arrow and e_0 differs from e_1 only by the small diode drop. During this particular quarter cycle most of the current passed by the rectifier is spent in charging capacitor C.

After b the input voltage starts to drop, e_0 drops also, and C discharges. However, the rectifier cannot conduct in the reverse direction, which forces C to discharge through R_L. Therefore voltage e_0 can only drop off exponentially at a rate determined by the values of C and R_L. Line bdf shows this discharge. (Appendix A contains a discussion of the exponential R-C discharge.)

The input voltage continues to follow the sine curve and for a considerable period of time remains below the value of e_0. No conduction can take place during this period because the diode voltage is reversed. At point d the input voltage again attempts to exceed the output voltage, and for a brief period C recharges. At point e the diode again stops conducting, and the process is repeated cycle after cycle.

The diode current consists of an initial large pulse followed by a series of equal charging pulses. The output current, on the other hand, equals e_0/R_L and remains relatively smooth. Thus the action of the circuit is much like that of a bucket with a small hole in the bottom through which water trickles constantly although the bucket is refilled periodically with a dipper. Since, except for the initial pulse, the average current through the capacitor equals zero, the average value of i_1 must equal that of i_0. Consequently, the peak diode current greatly exceeds the output current; for example, if the diode conducts 20 percent of the time, the pulse peak is more than five times the output current. This disadvantage of the circuit precludes its use for handling large amounts of power. For small currents, however, where even a small diode has more than sufficient current-carrying capacity, the simplicity of the circuit outweighs its disadvantages.

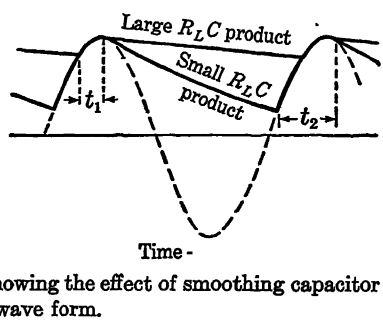


FIG. 3.2. Curves showing the effect of smoothing capacitor C and load resistor R_L on the output wave form.

The effect on the output voltage of varying C or R_L is shown by Fig. 3.2. The rate at which C discharges depends on the product

CR_L; a large C, or R_L, or both will make the circuit discharge slowly. In this case the output will be smooth with a short conduction period t_1 causing a high ratio of peak to average current. A low C or high load current (low R_L) will result in poor smoothing, longer conduction period t_2, and lower ratio of peak to average current.

It is important to observe that with a large CR_L product the output voltage approaches the peak value of the a-c input. Thus with an input of 100 volts effective value, the d-c output can approach 141 volts. Some circuits used in measuring equipment are especially designed to give this type of operation. This arrangement is also commonly used in radio receivers for demodulating the radio-frequency wave.

Another important design factor is the peak inverse voltage that appears across the rectifying element during that period in which it does not conduct. Figure 3.1 shows this value marked *PIV*. It approaches twice the peak applied voltage with low load currents that permit the output voltage to approach the peak input voltage.

3.2 Full-wave Circuit with Smoothing Capacitor

The circuit of Fig. 3.3 differs from that of Fig. 3.1 in the use of a center-tapped transformer and a second diode. The operation is similar to the half-wave circuit except that the diodes conduct alternately and charge the capacitor twice each cycle. This produces less output ripple and a ripple frequency of twice the incoming frequency. Also, with two pulses per cycle, the peak diode current is less than in the half-wave circuit for a corresponding output current, but the need for two diodes somewhat offsets this advantage. The wave forms of Fig. 3.3 show a peak inverse voltage nearly equal to twice the peak a-c voltage applied to *one diode.*

Although the full-wave circuit requires twice the number of secondary transformer turns to produce a given output voltage, this is balanced by the fact that the effective current drawn from each half is less than in the single-phase circuit. In fact, the advantages of good smoothing and doubled ripple frequency are so pronounced that circuits of this type provide the direct current required for the operation of the majority of radios and other small pieces of electronic equipment.

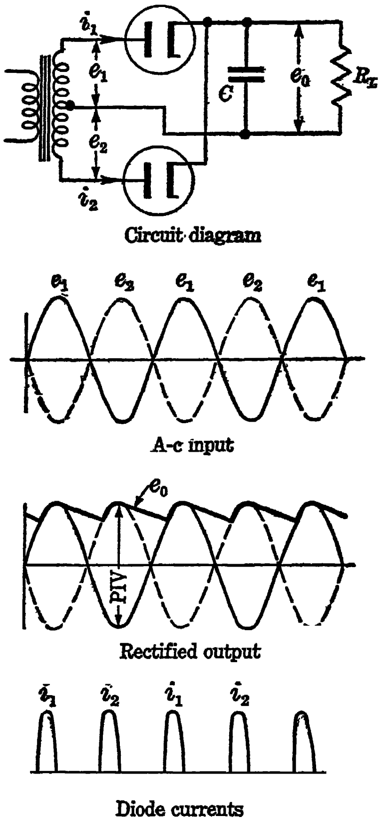


FIG. 3.3. Circuit diagram and wave forms for a full-wave rectifier with smoothing capacitor.

3.3 Full-wave Circuit with Resistance Load

Removing the capacitor from the circuit of Fig. 3.3 changes the operation considerably. As shown by Fig. 3.4 the diodes conduct alternately for full half cycles, making the output voltage approximately a series of half sine waves. For some purposes this rough output is useful; for instance, with a low-voltage transformer and suitable contact rectifiers, the circuit might be used to charge batteries. However, the main reason for studying this elementary circuit here is to provide a basis for extending the analysis to more complex circuits with smoothing filters.

For a given transformer voltage the peak inverse diode voltage

is the same as that shown in Fig. 3.3, but the average output voltage obtained (called the d-c component) is considerably less than the peak alternating voltage. In fact, computing the average value of a half sine wave shows the d-c output voltage to be only $2/\pi$ times the peak value. This makes the *effective* value of the

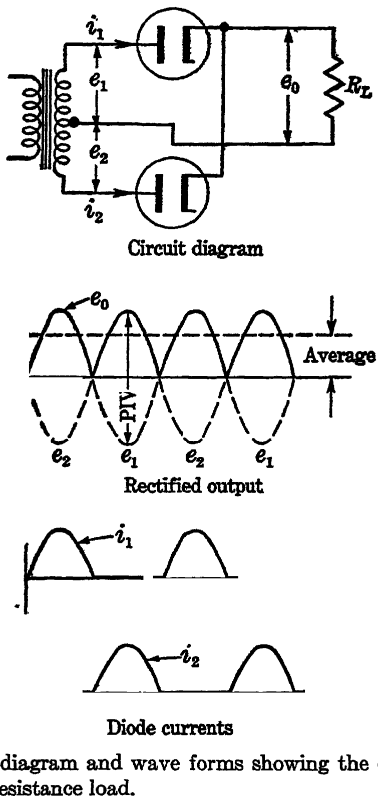


FIG. 3.4. Circuit diagram and wave forms showing the operation of a full-wave circuit with resistance load.

alternating voltage applied to one diode equal to 1.1 times the d-c output in an ideal circuit with no tube voltage drop. The factor 1.1 is the ratio of the effective ($0.707E_{max}$) to the average ($2/\pi E_{max}$) value for a half sine wave. (See Appendix B.)

3.4 Full-wave Bridge Circuit

The full-wave bridge circuit of Fig. 3.5 produces the same result as the circuit of Fig. 3.4, but with a simpler transformer and four rectifying elements instead of two. The diagram also shows the

symbol for contact rectifiers instead of diodes because the bridge circuit is particularly well adapted to the low forward voltage drop and limited inverse peak voltage capacity of a contact rectifier. The arrowhead and bar symbol shows the forward direction of current flow in the rectifying element; no appreciable current can flow in the opposite direction.

During the half cycle in which the transformer has the polarity

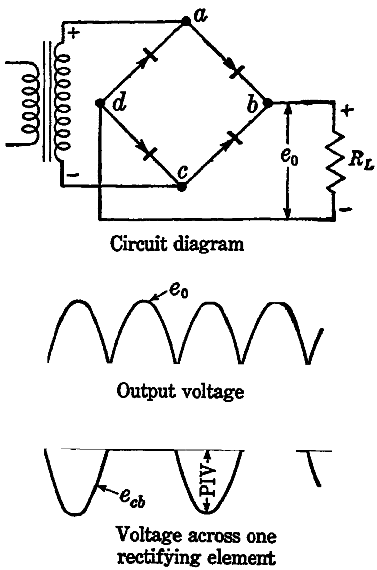


FIG. 3.5. The full-wave bridge rectifier with resistance load.

shown on the diagram, the current flows from a to b, through the load resistor, on to d, and returns to the transformer through point c. Although the current apparently has two choices of path when it reaches point d, it cannot travel to a but must go to the point of lower potential, c. For one half cycle the voltage across the load resistor equals the transformer voltage less the drop in *two* rectifying elements. On the reverse half cycle the current follows path *cbda* which makes the load current have the *same* direction. The net result is the same as though the transformer connections to the load resistor were reversed each half cycle, and the output voltage looks like a series of half sine waves.

The inverse voltage across one rectifying element, as shown by the curve of e_{cb}, has a maximum value approaching the peak of

the applied alternating voltage. This occurs for the rectifying element between b and c during the peak of the positive half cycle when element ab conducts and practically the full transformer voltage appears between b and c. Comparison with Fig. 3.4 shows the bridge circuit to have only half the peak inverse voltage for the same d-c output. For this reason the bridge circuit is used for very high-voltage rectifiers, where the chief limitation is the design of tubes for high inverse voltage, and for circuits employing contact rectifiers with limited inverse peak ratings.

A further advantage of the circuit is increased transformer utilization. Instead of a double winding, each part of which operates only half the time, the bridge rectifier uses a single winding operating on both halves of the cycle. The chief disadvantages is the need for four rectifying elements and the fact that two elements always appear in series with the load. This doubles the voltage drop in the rectifier and decreases the efficiency.

3.5 Filters or Smoothing Circuits

Vacuum-tube amplifiers, radio receivers, and many other pieces of apparatus require a much smoother source of direct current than any of the circuits so far described. To provide this smooth-

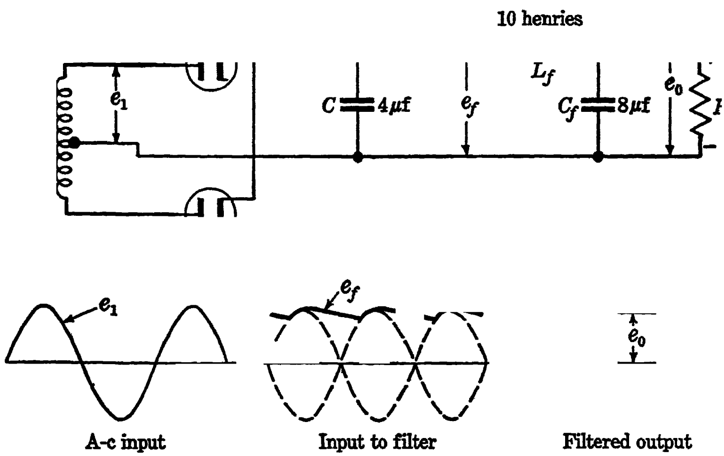


FIG. 3.6. Full-wave rectifier with capacitance-input (π-section) filter.

ing, additional circuits called *filters* are added to the output of the rectifier. Figures 3.6 and 3.7 show two examples of filters added to full-wave rectifier circuits. Figure 3.6 can be thought of as a

section of filtering added to the circuit of Fig. 3.3, while Fig. 3.7 represents a filter added to the circuit of Fig. 3.4. Although the common name of condenser input or π-section filter suggests that C is considered as part of the filter in Fig. 3.6, it is more instructive to think of the filter as composed of L_f and C_f alone, while C remains part of the basic rectifier circuit.

The addition of a filter between the rectifier and the load does

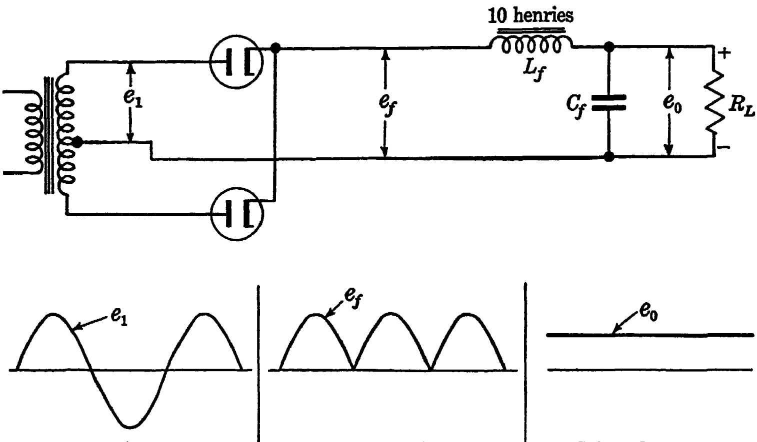


FIG. 3.7. Full-wave rectifier with inductance-input (L-section) filter.

change the current drawn from the rectifier. An inductance produces a counter-emf tending to resist any current *change* through it and thus smooths out the current taken from the rectifying circuit. This slightly straightens out the voltage-decay curve across C in Fig. 3.6, but the change is inconsequential and the diode current pulses remain about the same.

In the circuit of Fig. 3.7, however, the current passing through L_f equals the current delivered by the rectifying elements, and the inductance has the important property of smoothing out this current and reducing the ratio of peak to average value. Figure 3.8 illustrates this effect and shows that with sufficient inductance the ratio of peak to average current can approach unity. The filter input voltage remains essentially a series of half sine waves because of the low voltage drop in the rectifying elements.

The filter inductance L_f must have a minimum of resistance to make the direct voltage drop small compared with the output vol-

tage. However, it must have a high *reactance* to the *ripple* components of the filter input voltage.

Harmonic Series. An extremely useful approach to the smoothing problem is to consider the filter input voltage as the sum of a steady direct voltage and a number of sinusoidal harmonics. This point of view is supported by the Fourier series of mathematics which states that any single-valued repeating function can be

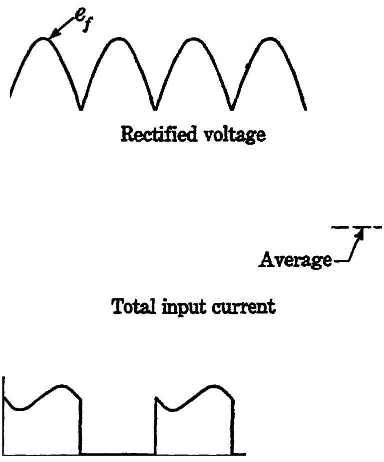


FIG. 3.8. Current and voltage wave forms at the filter input of Fig. 3.7.

represented as a harmonic series of sine waves. By harmonic series we mean that the frequencies of the component terms are integral multiples of the fundamental repetition frequency of the wave.

To provide a graphic illustration of this series representation, Fig. 3.9 shows the synthesis of a complex wave from its harmonic components. The components shown are a constant of three units height, a fundamental sine wave with a five-unit peak, and a second harmonic of two units peak height that lags 90 degrees. Directly below these waves the diagram shows the sum of these three components, which gives a complex wave having the equation

$$y = 3 + 5 \sin \omega t + 2 \sin\left(\omega t - \frac{\pi}{2}\right)$$

This complex wave repeats its form each cycle of the fundamental because at the beginning of each cycle the second harmonic is

again in the same relative position. Had the frequency ratio been not integral, the wave could not have repeated.

Of course we are primarily interested in the reverse process of *analyzing* a complex wave, that is, starting with a given wave and finding its various harmonic components. This can be done for

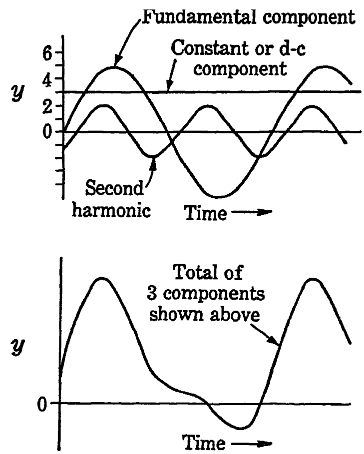


FIG. 3.9. Graphical synthesis of a complex wave having a shape expressed by the equation $y = 3 + 5 \sin \omega t + 2 \sin (\omega t - \pi/2)$.

any arbitrary wave shape, and it is a particularly easy process for the relatively simple output of a full-wave rectifier. The result is

$$e = 0.636E_m + 0.424E_m \cos \omega t - 0.085E_m \cos 2\omega t + 0.036E_m \cos 3\omega t + \cdots \quad (3.1)$$

where E_m is the peak height of the original rectified wave shape. The series has an infinite number of terms, but the magnitudes drop so rapidly with increasing order that relatively few of them are necessary to represent the actual wave with accuracy. The first term of Eq. (3.1) represents the average or d-c component because each of the remaining cosine terms has an average of zero.

Figure 3.10 shows the first four components of Eq. (3.1) in graphic form together with the original rectified wave being analyzed and, as a matter of interest, a number of points showing the sum of the four components. The difference between these points and the original wave represents the error due to neglecting the remainder of the infinite series of harmonics. This diagram also shows the waves on a scale chosen to make the d-c component 100 units high; the labels on the alternating components thus

show their *effective* values in percent of the d-c component for convenience in later computations.

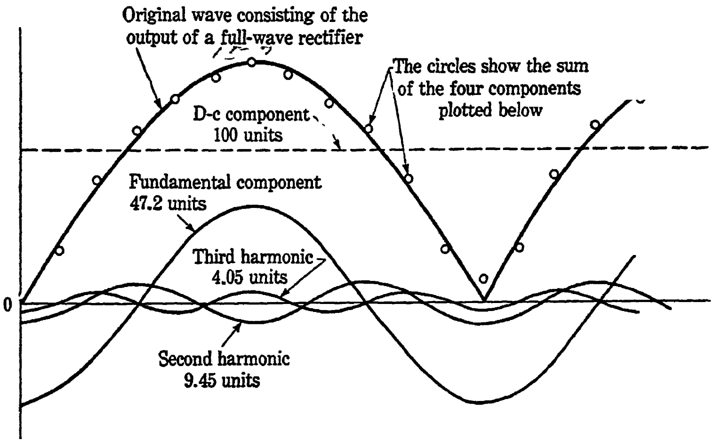
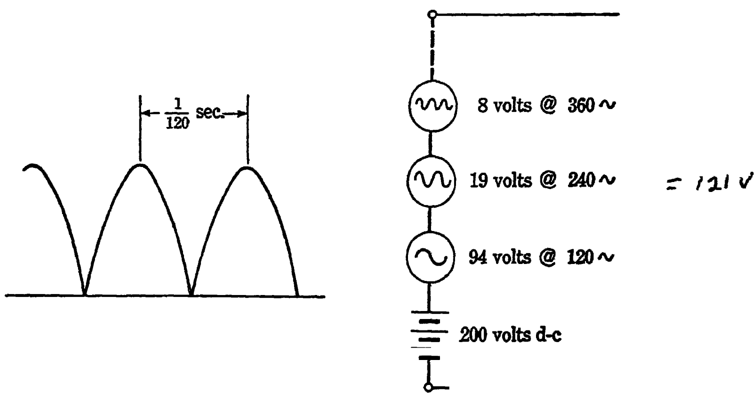


FIG. 3.10. This illustrates the breakdown of the output wave of a full-wave rectifier into a d-c component and a number of harmonics. The scale is chosen to make the d-c component 100 units high.



(a) Rectified voltage (b) Equivalent series generators

FIG. 3.11. Equivalent representations of a full-wave rectifier output having an average value of 200 volts.

In the following discussion of the smoothing action of a filter we shall use the information of Fig. 3.10 together with the *superposition* principle of electrical circuits. This amounts to thinking of the rectified voltage as the series of generators shown in Fig. 3.11,

then computing the filter output voltage for each individual component of the input voltage, and finally adding up the individual output components to obtain the total output voltage.

Smoothing Factor α. We shall now investigate the behavior of the filter for one particular ripple frequency. By rearranging the filter as shown in Fig. 3.12, it can be represented as a voltage divider in which E_1 represents one component of the input ripple (say the 94 volts of 120 cycles) and E_r is the ripple voltage of the same frequency appearing across the load. The ratio of E_1 to E_r

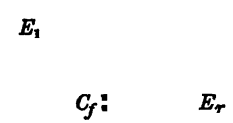


FIG. 3.12. Analysis of an L-section filter.

will be defined as the *smoothing factor* α. Good smoothing requires a large smoothing factor; to make it large the series impedance of L_f must greatly exceed the parallel impedance of R_L and C_f. Since R_L is not necessarily small, the reactance of C_f must be, and as a safe approximation we shall assume that the parallel impedance is due to C_f alone. We shall also neglect the resistance of the choke L_f because at ripple frequencies the reactance of a practical iron-cored inductor is large compared with its resistance.

With these approximations the alternating current flowing through L_f and C_f equals $E_1/(X_L - X_c)$. This current can produce a voltage drop in the reactance X_c of

$$E_r = IX_c = \frac{E_1(X_c)}{X_L - X_c}$$

Solving for the ratio E_1/E_r, we obtain

$$\alpha = \frac{E_1}{E_r} = \frac{X_L - X_c}{X_c} = \frac{X_L}{X_c} - 1$$

Replacing X_L by ωL and X_c by $1/\omega C$, we get

$$\alpha = \omega^2 LC - 1 = (2\pi f)^2 LC - 1 \quad = \frac{E_f}{E_r} \tag{3.2}$$

As an example, the inductance of 10 henrys and capacitance of 8 microfarads shown in Fig. 3.6 gives a smoothing factor of 44 for 120 cycles. This means that the filter reduces the 120-cycle component of the input voltage by a factor of 44 before it reaches the output. The smoothing factor for the 240-cycle component is nearly four times as great, and for still higher frequency components the smoothing factor jumps to even higher values. As a result the filter output contains very little ripple.

TABLE 3.1. SUMMARY OF FILTER OPERATION

Component	Frequency	Input, volts	Smoothing factor	Output, volts
D-c..............	0	200	...	188.00*
Fundamental.....	120	94	44	2.20
2d harmonic......	240	19	180	0.10
3d harmonic......	360	8	410	0.02

* The voltage loss in the choke resistance accounts for the drop in d-c output. This figure was computed for a current of 80 ma passing through a choke having 150 ohms of resistance.

Table 3.1 clearly shows that the only appreciable output ripple is the fundamental component. Consequently, the computation of filter performance resolves itself into calculation of the d-c drop in the choke and determination of the smoothing factor for the fundamental ripple frequency alone.

The R-C Filter. Good smoothing can be obtained with a filter section in which a resistance replaces the inductance (Fig. 3.13).

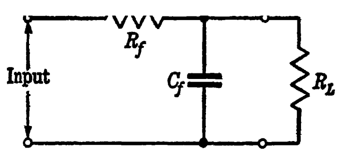


FIG. 3.13. A resistance-capacitance (R-C) filter.

This has the disadvantage of causing considerable loss of d-c output, but in many cases the additional filtering is needed for a load drawing only a few milliamperes at a low voltage.

Following the previous analysis with similar simplifying as-

sumptions, we obtain

$$E_r = IX_c = \frac{E_1 X_c}{R_f}$$

This rearranged to solve for α gives

$$\frac{E_1}{E_r} = \frac{R_f}{X_c} = \omega R_f C_f \tag{3.3}$$

As a practical example we shall imagine that the output of an L-C filter consists of 300 volts (d-c) with 1 volt of 120-cycle ripple. Most of the output current is used at this voltage, but one part of the load circuit draws only 2 milliamperes at 250 volts and requires that the ripple be less than 15 millivolts. To meet these requirements, resistor R_f must carry the 2 milliamperes with a drop of 50 volts. This makes its value

$$R_f = \frac{50}{0.002} = 25{,}000 \text{ ohms}$$

The smoothing factor for the R-C section equals the ratio of input to output ripple.

$$\alpha = \frac{1}{0.015} \quad 67$$

By placing the known values of α and R_f in Eq. (3.3) we can obtain the value of C_f.

$$C_f = \frac{67}{(2\pi)(120)(25{,}000)} = 3.5 \times 10^{-6} \text{ farad}$$

The next *larger* commercial condenser size should be chosen because *too little* ripple is always tolerable.

Figure 3.14 shows the complete two-stage filter. This represents

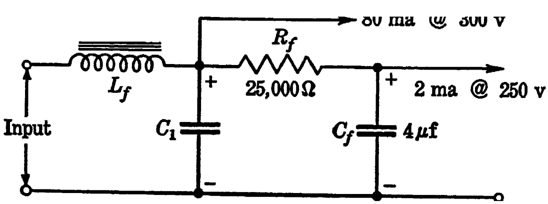


FIG. 3.14. A two-section filter.

good engineering practice of providing just sufficient smoothing for each load at a minimum expense.

3.6 Rectifier Regulation

The term *regulation* refers to the variation of terminal voltage when load current is drawn from a power source. Good regulation indicates that the terminal voltage remains essentially constant with load variations; the output voltage of a circuit with poor regulation drops rapidly with increasing load current.

The typical regulation curves of Fig. 3.15 show that rectifier circuits with initial smoothing capacitors have poor regulation. A study of Fig. 3.6 to find the cause of this poor regulation indicates that the load current causes voltage drops because of

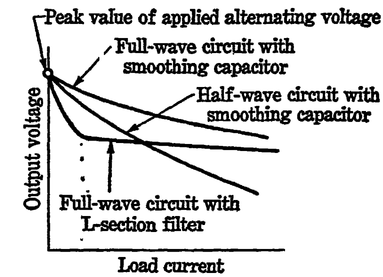


FIG. 3.15. Typical rectifier regulation curves.

'CRITACAL pr. - CRITICA

(1) transformer winding resistance, (2) drop in the rectifying elements, (3) discharge of C between periods of recharge, and (4) drop in the resistance of L_f. Item 1 is small because transformers are efficient devices. The second voltage loss depends upon the rectifying element chosen, but in any case the presence of C causes the current to flow in pulses of high peak value and correspondingly large voltage loss in the rectifier and transformer. An even more important source of drop is the capacitor discharge between cycles. At low load currents the capacitor discharges slowly and the output voltage approaches the peak value of the input voltage, as shown by Fig. 3.2. At high loads (low R_L) the capacitor discharges more, as shown by the lower curve of Fig. 3.2. This reduces the average output voltage and contributes to poor regulation. The remaining voltage drop in the choke, L_f, depends upon the effort taken to reduce the choke resistance, and it can be made small with proper design.

The curve for the inductance input filter (circuit of Fig. 3.7) shows excellent regulation over most of the load range. The im-

provement is due to elimination of the discharge of capacitor C, as discussed above.

In addition, the lower peak current demanded of the rectifier minimizes the voltage drop in the transformer and rectifying element. In fact, with the use of gas-filled tubes having substantially constant voltage drop, and with generous design of transformer and choke, the regulation curve can be made substantially flat over most of the range.

The sudden rise at low currents on the inductance-input regulation curve is caused by the inability of L_f to maintain continuous current flow. At loads below this critical point the operation tends to become more like that of a capacitor-input circuit and the voltage approaches the peak value of the a-c input. Poor regulation can be avoided by always operating at loads above the critical value. For this purpose a fixed resistor called a bleeder is sometimes connected to the output terminals.

PROBLEMS

3.1 Show that the average value of a half sine-wave loop is $2/\pi$ times the maximum value.

3.2 A simple single-phase bridge rectifier without filter provides an average output voltage (d-c component) of 10 volts across a 5-ohm load resistor. Assuming negligible drop in the rectifying elements, compute (*a*) the required transformer voltage, (*b*) the peak current carried by a rectifying element, and (*c*) the peak inverse voltage across one element.

3.3 The shunt diode rectifier of Fig. 3.16 operates with a very high load

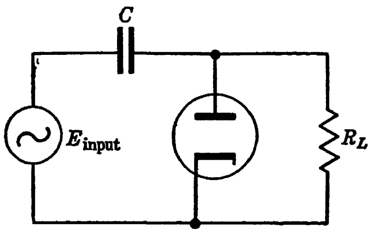


FIG. 3.16. The shunt diode rectifier.

resistance so that the rectifier output voltage approaches the peak input voltage. The circuit has a very long time constant compared with the length of 1 cycle, and the diode operates essentially as an ideal conductor in the forward direction. Develop wave forms of the output voltage and the voltage across C. Mark the polarities of each voltage on the diagram.

3.4 The rectifier of Fig. 3.6 provides a filter input voltage of 300 volts (d-c component) plus a fundamental a-c component of 40 volts and numerous

other alternating components of higher frequencies. The supply-line frequency is 60 cycles, the choke has 200 ohms resistance, and the load draws 120 ma. Compute (*a*) the direct output voltage. (*b*) the effective output ripple voltage, and (*c*) the percent ripple as compared with the d-c component.

3.5 A rectifier having the circuit of Fig. 3.7 operates from a 400-cycle aircraft supply and provides a filtered output voltage of 500 volts at a load current of 300 ma. The 1-henry choke has a resistance of 150 ohms, and the output voltage must not contain more than 0.5 percent ripple. Determine the minimum size of C_f.

3.6 Two 10-henry inductors and two 5-μf capacitors are available for filtering the output of a 60-cycle full-wave rectifier. Is it better to connect them into a single L section of 20 henrys and 10 μf, or into a double section filter? Make computations to substantiate the choice.

3.7 The input to the two-section filter of Fig. 3.14 comes from a full-wave rectifier operating from a 60-cycle supply. The choke has a resistance of 250 ohms and an inductance of 15 henrys, and C_1 equals 10 μf. Compute the ripple voltage in the 250-volt output.

F.

CHAPTER 4

GRID-CONTROLLED VACUUM TUBES

ALTHOUGH the development of the diode provided many interesting possibilities, it was the invention of the triode by De Forest that really opened the way for the development of modern electronics. The long-distance telephone, radio, television, electronic control and measuring instruments, radar, and countless other devices all depend on the basic process of amplification provided by this tube.

4.1 The Triode

The idea behind the triode is deceptively simple. Why not place between the anode and cathode of a diode a third electrode consisting of a grid of wires to control the flow of electrons to the plate? By making this grid negative, the effect of the space charge in repelling the slower electrons to the cathode can be increased and the plate current decreased. With sufficient negative grid voltage, the plate current can even be shut off. This provides us with an electrical valve for controlling the current

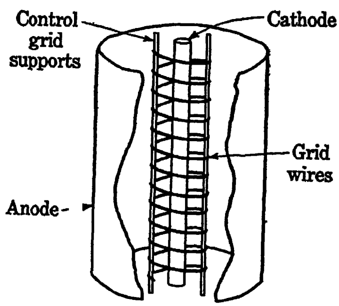


FIG. 4-1. Cutaway view of a triode.

flowing through the tube—a much more delicate control than the diode provides. In addition, the plate current is controlled with

the expenditure of almost no control power, similar to a well-greased valve that can be turned by a small boy to release a flow of water representing a relatively large amount of power and perhaps damage.

Figure 4.1 shows a cutaway view of a small triode. The spiral grid surrounding the cylindrical cathode is supported by two heavier vertical wires with each intersection spot-welded for rigidity. The anode in turn surrounds this assembly. The tube characteristics are controlled by the grid wire size and spacing and by the relative grid and plate diameters. Not all tubes are constructed with concentric circular elements; they are often elliptical or rectangular in section. Figure 4.2 shows the conventional symbol for a triode.

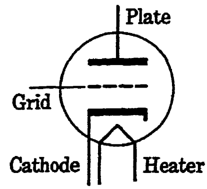


FIG. 4.2. Conventional diagram for a triode.

4.2 Triode Characteristics

To understand the characteristics of a triode we shall perform an imaginary experiment using the circuit of Fig. 4.3. As a simplification the heater connections are not shown, and it is to be understood that the tube operates at rated cathode temperature. Under this condition the emission is greater than needed, and saturation will not be reached. This is the only condition of practical interest to us.

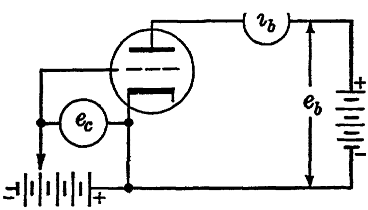


FIG. 4.3. A circuit for determining the triode characteristic curves. All voltages are measured with respect to the cathode.

If the plate voltage e_b is held constant at 100 volts and the grid voltage varied, the heavy curve of Fig. 4.4 results. With zero grid voltage the anode current is 10.4 milliamperes, but making the grid more negative assists in repelling electrons back to the cathode and the current drops off. A grid voltage e_c of about −6 volts reduces the plate current to zero. This is called the point

of *cutoff*. No curve of grid current is shown because a negative grid repels electrons and there is practically no current flow to it. Since power equals the product of voltage and current, no power is taken from the source of grid voltage as long as the grid is negative with respect to the cathode. For positive grid voltages the plate current continues to rise in a smooth curve, but the grid

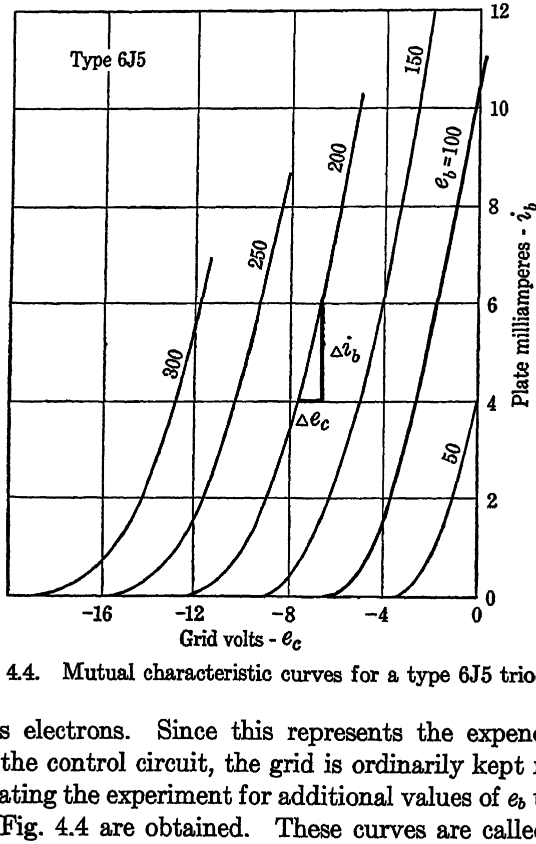


FIG. 4.4. Mutual characteristic curves for a type 6J5 triode.

also draws electrons. Since this represents the expenditure of power by the control circuit, the grid is ordinarily kept negative.

By repeating the experiment for additional values of e_b the other curves of Fig. 4.4 are obtained. These curves are called *mutual* characteristics because they show the effect of a voltage in one circuit (the grid) on the current in a second circuit (the plate). The curves are not straight, although it would usually be desirable to have them so.

Transconductance. The curve slope measures the effectiveness of the grid in producing plate-current changes—the steeper the

curve, the better. This slope is called the transconductance, defined mathematically by the expression

$$g_m = \frac{\partial i_b}{\partial e_c} \qquad \text{mhos} \tag{4.1}$$

The partial derivative merely means that the plate voltage is held constant while measuring the ratio.

Transconductance can be obtained by drawing a tangent to one of the curves of Fig. 4.4 and measuring the slope in milliamperes per volt, or by taking small incremental changes and finding g_m from the ratio

$$g_m = \left.\frac{\Delta i_b}{\Delta e_c}\right|_{e_b = \text{constant}} \tag{4.2}$$

A sample determination of g_m at a plate voltage of 200 volts and a plate current of 5 milliamperes is indicated by the triangle drawn on Fig. 4.4. Reading the values of Δi_b and Δe_c from the triangle and placing them into Eq. (4.2)

$$g_m = \frac{0.002}{1} = 0.002 \text{ mho or } 2{,}000\ \mu\text{mho}$$

At normal plate currents most tubes have transconductances of the order of 2,000 micromhos.

Inspection of the curves will show g_m to be essentially a function of the plate current; for a given current all the curves have about the same slope.

Plate Characteristics. A more useful set of characteristics can be obtained by replotting the data of Fig. 4.4 to show plate current i_b as a function of plate voltage e_b. Figure 4.5, which is taken from a standard radio-tube-design manual, shows a complete family of such curves for a type 6J5 triode. This is called a set of *plate* characteristic curves because the ordinate and abscissa represent current and voltage in the plate circuit only. Although the curves of Fig. 4.4 are often useful for discussing the operation of a circuit, the curves of Fig. 4.5 are more convenient for actually computing the performance of the tube and its associated circuit.

Plate Resistance. Inspection of the plate characteristics shows them to be curved and nearly equally spaced horizontally. For a given plate current, the curves have about the same slope. Since

a tube operating as an amplifier experiences continuous alternating changes in plate voltage and plate current, it is important to have a measure of the ratio between a small plate-voltage change and the corresponding plate-current change. This ratio is called the dynamic or plate resistance, defined as

$$r_p = \frac{\partial e_b}{\partial i_b} \quad \text{ohms} \tag{4.3}$$

In terms of finite incremental changes this becomes

$$r_p = \frac{\Delta e_b}{\Delta i_b}\bigg|_{e_c = \text{constant}} \tag{4.4}$$

This factor has the dimensions of volts per ampere, or ohms, and it is called the *plate* resistance because it deals with the plate circuit only.

On the plate characteristic curves r_p is represented by the inverse of the slope. Thus at high plate currents the plate resistance is low; a small change in voltage produces a large change in current. At low currents the plate resistance is higher, and as the current is reduced to zero the plate resistance approaches extremely high values.

The triangle *abc* drawn on Fig. 4.5 provides the data required for computing the plate resistance at a grid voltage of -10 volts and a plate current of 7 milliamperes. Reading the values of Δi_b and Δe_b from the triangle and placing them into Eq. (4.4)

$$r_p = \frac{20}{0.002} = 10{,}000 \text{ ohms}$$

Amplification Factor. The concept of amplification factor is an important one; it might be called the electrical advantage of the tube in comparison to the mechanical advantage of a simple lever. One way of measuring the mechanical advantage of a lever, as suggested by Fig. 4.6, would be to apply two forces of such size as to produce no deflection of the arm. The ratio of the two forces gives the mechanical advantage of the system.

In a somewhat similar fashion the amplification factor is defined as the ratio of a change in plate voltage to a change in grid voltage that will produce no change in plate current. Mathematically

$$\mu = -\frac{\partial e_b}{\partial e_c} \tag{4.5}$$

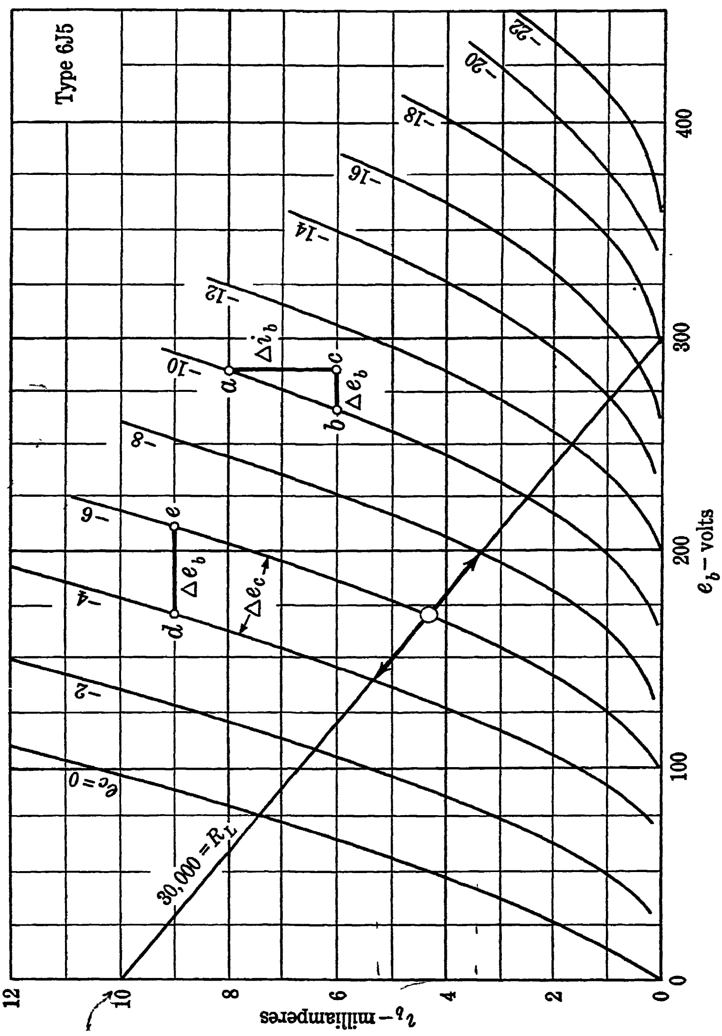


FIG. 4.5. 'amily of 6J5 iode plate racte

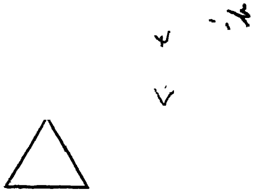


FIG. 4.6. A method of measuring mechanical advantage. At balance the mechanical advantage equals F_1/F_2.

or expressed as finite incremental changes

$$\mu = -\frac{\Delta e_b}{\Delta e_c}\bigg|_{i_b = \text{constant}} \tag{4.6}$$

The minus sign is used to make factor μ come out positive. This can best be illustrated by referring to the horizontal line *de* drawn on Fig. 4.5. The line is horizontal to represent a constant plate current. Moving from *d* to *e* involves a change of *plus* 40 volts in plate voltage and a change of *minus* 2 volts in grid voltage (from the −4 line to the −6 line). Using these increments

$$\mu = \frac{-40}{-2} = +20$$

The amplification factor is a pure number because it is the ratio between two voltages.

Since the curves of Fig. 4.5 show equal increments of grid voltage, the horizontal spacing between curves (corresponding to line *de*) is directly proportional to the factor μ. By measuring this spacing at various points on the curves it will be found that the amplification factor is surprisingly constant except at very low currents. In fact μ is often called a triode constant—a term not justified for the plate resistance and the transconductance.

A tube with an amplification factor of less than about 10 is called a low-mu triode; medium mu applies to those between 10 and 30; and high-mu tubes have amplification factors between 30 and 100.

Relation between Factors. The relation between the three factors can be determined with the help of Fig. 4.7, which shows a small section of the characteristic curves of Fig. 4.5. To determine r_p we hold the grid voltage constant and move along the line from *a* to *c*. This involves a change in plate voltage from *a* to *b* (written *ab*) and the corresponding change in plate current *bc*. Then r_p is computed from the ratio *ab*/*bc*. To measure g_m we must hold the plate voltage constant, which means moving along the vertical line *bc*. This involves a plate-current change *bc* and the corresponding grid-voltage change from *e* to *d* (*ed*). Notice that the lengths *ab* and *bc* represent actual voltage and current changes, whereas *ed* represents the difference between the grid voltages marked on the two curves ($K_1 - K_2$).

The amplification factor is measured by the ratio of plate-voltage change *ab* to grid-voltage change *de*. Observe that here the direction of grid-voltage change is reversed from that used in the determination of g_m (*de* is the negative of *ed*). Using this fact and comparing the ratios as shown in Fig. 4.7, we find

$$\mu = g_m r_p \tag{4.7}$$

Hence the determination of any two of the three factors is sufficient to define the other. Factors μ and r_p are of greatest significance

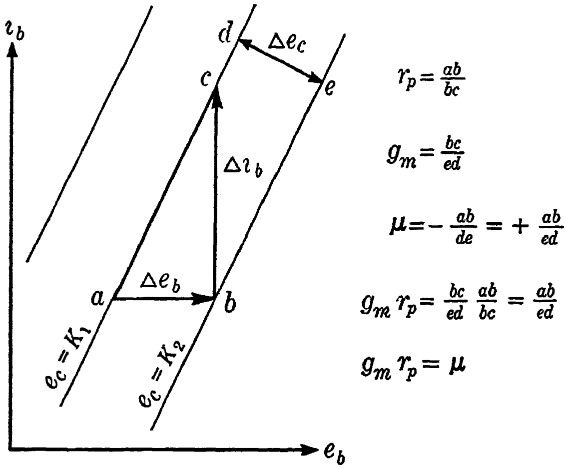


FIG. 4.7. A graphical illustration of the relationship between the triode coefficients.

with triodes, but the transconductance is especially convenient for describing the action of pentodes.

4.3 Voltage Amplification

One important application of the triode is to obtain voltage amplification. By this is meant the ability of a circuit to take a small alternating voltage from some source such as a microphone and produce a much larger output voltage that accurately follows every detail of the input. In one sense a step-up transformer performs this operation, but the transformer demands that the signal provide even more input power than output power, whereas the vacuum-tube circuit draws no power from the input but obtains the output power from some other source such as a battery.

Figure 4.8 shows the circuit diagram for a simple triode amplifier capable of amplifying the signal voltage e_g. The circle drawn for e_g represents some source of signal voltage to be amplified (micro-

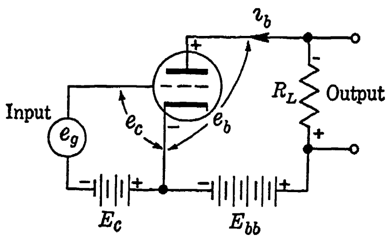


FIG. 4.8. Circuit diagram for a simple triode amplifier.

phone, phonograph pickup, strain gage, etc.). Battery E_c is placed in the circuit to keep the grid negative at all times regardless of the polarity of e_g. This is called the *bias* battery, and E_c is called the *grid-bias* voltage. If, for instance, E_c is -6 volts, the maximum allowable value of e_g would be $+6$ volts without driving the grid positive. The negative grid prevents grid-current flow so that the source of input voltage provides no power.

In the plate circuit, battery E_{bb} (*bb* for "B" battery) acts as the source of power to maintain the anode positive. Load resistor R_L is placed in the circuit for the purpose of converting the plate-current variations into voltage variations. For the purpose of amplification a mere change in *current* is of little use; it is variations in output *voltage* that are desired. This amplified signal voltage can then be used as the input for the next stage of amplification, etc.

Figure 4.9 shows a graphical construction illustrating the performance of the triode of Fig. 4.8. The analysis starts with a curve showing the relation between the plate current and grid voltage for the tube and its *associated circuit.* This curve is called the *dynamic characteristic*, and it is *not* one of the static curves of Fig. 4.4. The curve of Fig. 4.9 represents the relation between i_b and e_c with *variable* plate voltage because each value of plate current produces a different voltage drop in R_L. Let us assume, for the moment, that this dynamic characteristic has already been determined.

Just below the dynamic curve the figure shows a picture of the grid voltage, consisting of the constant negative value of E_c plus

a sine wave representing the incoming signal. Of course the signal voltage would ordinarily be a complex wave shape, but for the purpose of illustration a simple sine wave has been chosen. On this part of the diagram the vertical scale represents time, and the horizontal distances represent the grid voltage. With the diagram

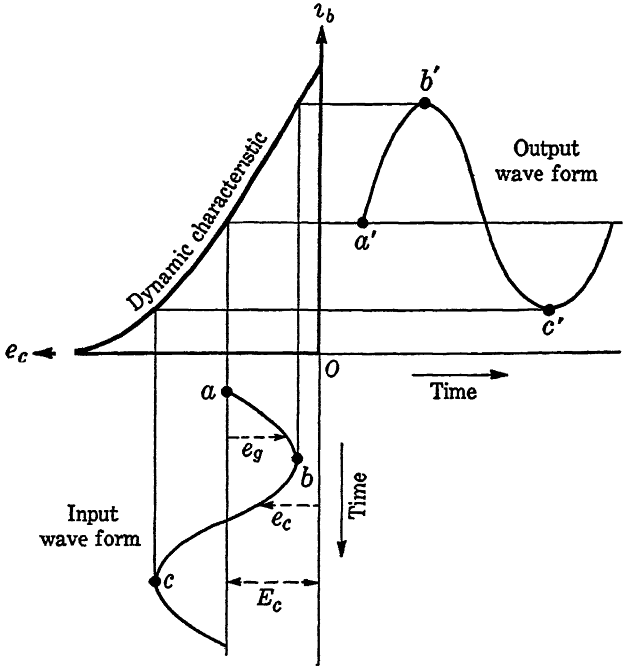


FIG. 4.9. Development of the plate-current wave form for a triode amplifier.

rotated a quarter turn counterclockwise the grid voltage wave will be seen with a more familiar orientation.

Since the dynamic characteristic shows the relation between grid voltage and plate current we can now take the grid voltage at any time moment and graphically find the corresponding plate current at that moment. For instance, point a on the grid-voltage curve represents a moment when e_g equals zero and the total grid voltage is just E_c. By running a vertical line to intersect the curve and then proceeding horizontally we obtain the point a' on the current wave. In the same fashion point b' is found from b, and so on until the plate-current wave has been developed.

If the dynamic characteristic were a straight line, the plate current would be accurately proportional to the grid voltage and the output wave would look exactly like the input wave form. The actual dynamic characteristic is curved, however, and the upper peak of the plate-current wave is larger than the lower peak even though the input wave was symmetrical. Thus the amplitude of the output is not exactly proportional to the input, and *amplitude* or *nonlinear* distortion is said to be present. Amplitude distortion is important in amplifiers operating with relatively large input signals.

For tiny signals, only a small portion of the dynamic curve is used, there is less curvature in this restricted length, and the amount of distortion is very small. Consequently, amplitude distortion is unimportant for small signals, say less than 1 volt. In the following discussion we shall assume that this condition obtains and that the plate-current variations are accurate sine waves.

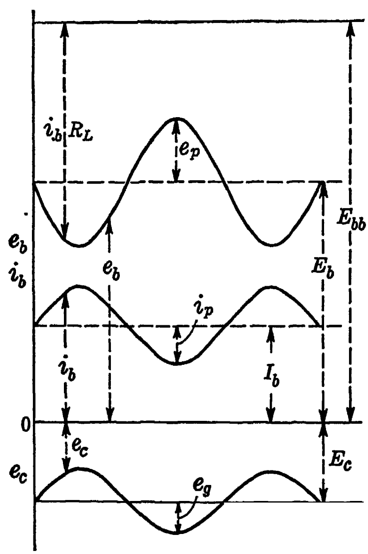


Fig. 4.10. Voltage and current wave forms for the triode amplifier of Fig. 4.8.

Figure 4.10 shows the wave forms and standard nomenclature for the amplifier circuit of Fig. 4.8 plotted on a common set of axes. The bottom wave form shows the grid voltage applied to

the tube and represents the input wave form to be amplified. Above this is plotted the plate-current wave form derived from the grid-voltage wave with the aid of Fig. 4.9.

With these waves before us let us stop for a moment to consider the meanings of the symbols used to describe them. The grid voltage comes from two sources of voltage applied to the circuit, E_c the fixed negative bias, and e_g the alternating input. Voltage E_c is called the average or d-c component, and e_g is named the alternating component. The sum of the two gives the depressed sine wave whose value remains always negative. The plate-current wave is similar in shape to the grid voltage except that it is everywhere positive. Although this curve actually represents a *single* current, it is convenient to think of it as being made up of *two* components similar to the grid voltage. As shown on the diagram these two components are I_b, the average or d-c component, and i_p the alternating component. In equation form

$$i_b = I_b + i_p \tag{4.8}$$

Capital letters refer to the direct currents or voltages and to the effective values of the alternating ones. Small letters refer to instantaneous values of varying quantities. Observe that subscripts c and b apply to quantities measured from the zero line, while g (for grid) and p (for plate) apply only to the alternating components.

The curve for the plate voltage is obtained by subtracting the voltage drop in resistor R_L from the supply voltage E_{bb}. Thus

$$\begin{aligned} e_b &= E_{bb} - (I_b + i_p)R_L \\ &= (E_{bb} - I_bR_L) - i_pR_L \\ &= \quad E_b \quad + \quad e_p \end{aligned} \tag{4.9}$$

As with the plate current, it is convenient to consider the plate voltage as made up of two components, E_b the direct or average value, and e_p the alternating component. As shown by Eq. (4.9)

$$\underline{E_b = E_{bb} - I_bR_L} \tag{4.10}$$

and

$$e_p = -i_pR_L \tag{4.11}$$

The negative sign in Eq. (4.11) indicates that the alternating voltage in the plate circuit is opposite in phase to the alternating component of the plate current. This is easily verified by an examination of Fig. 4.10.

Of course the output alternating voltage is always accompanied by a large direct voltage, but by using condensers or transformers it is possible to block off the d-c component and obtain the alternating wave alone. Such means are discussed in Chap. 9, devoted to practical amplifier circuits.

4.4 Load-line Analysis

The preceding discussion shows in a general way how an amplifier operates, but it does not answer the question of how much the input voltage is amplified. Let us now quantitatively determine the amplification of a given tube and circuit, as shown by Fig. 4.11. The static curves of Fig. 4.5 apply to the 6J5 chosen for this circuit.

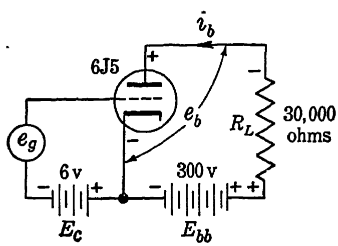


FIG. 4.11. Simple triode amplifier showing typical circuit constants.

At first glance, the problem of computing the alternating plate voltage e_p for a given grid voltage e_g with practically everything in the circuit varying seems somewhat formidable. However, the choice of supply voltage and load resistance immediately restricts the problem to a fairly simple one. From an inspection of Fig. 4.11,

$$e_b = E_{bb} - i_b R_L \tag{4.12}$$

Rearranging this to solve for i_b, we obtain

$$i_b = -\left(\frac{1}{R_L}\right) e_b + \frac{E_{bb}}{R_L} \tag{4.13}$$

This is of the form $y = mx + b$ and represents a straight line having an intercept on the i_b axis of E_{bb}/R_L, a slope of $-1/R_L$, and an intercept on the e_b axis of E_{bb}. This line, shown plotted on the $i_b - e_b$ curves of Fig. 4.5, is called the *load line*, and its position is determined entirely by the value of the supply voltage E_{bb} and by the load resistance. In this particular case the i_b-axis intercept equals 300/30,000, or 10 milliamperes, and the e_b-axis intercept is 300 volts.

Now we are able to determine the plate voltage for any particular grid voltage. For instance, if e_c is -6 volts at some moment,

the plate voltage at the same instant can be found from the junction of the -6 grid-voltage line and the load line. This is the only point on the chart that will simultaneously satisfy the tube characteristics and the remainder of the circuit. From the chart, the plate voltage is found to be 170 volts for $e_c = -6$.

Let us suppose that an alternating voltage e_g having a crest value of 2 volts is applied to the input. The grid voltage e_c will swing between the limits of -4 and -8 volts, and the point of operation will travel up and down the load line between the limits of the -4 and -8 curves, as shown by the heavy portion of the line. At the limits of the excursion the plate voltages read from the curve are 140 volts and 200 volts. Thus a 4-volt swing in the grid circuit produces a 60-volt swing in the plate circuit. In this particular case the circuit has a *voltage amplification* of 15 because a small voltage change in the grid circuit produces a change 15 times as large in the plate circuit. Expressing this relationship mathematically, the amplification is

$$A = \frac{e_p}{e_g} = \frac{E_p}{E_g} \tag{4.14}$$

The first ratio represents the ratio of the grid- and plate-voltage changes in general, while E_p/E_g represents the ratio of the effective values if the signal is a sine wave.

Do not confuse amplification with amplification factor. Amplification is the ratio between the output- and input-voltage changes for an actual circuit in which the plate current also changes. Amplification factor was defined as the ratio between plate- and grid-voltage changes for *constant* plate current. Although the actual amplification A of the *circuit* is 15, the amplification factor μ of the *tube* is 20. It is generally true that the amplification must always be less than the amplification factor.

An investigation of the load-line intercepts with the characteristic curves shows them to be nonuniform; at the lower end of the line the plate-voltage changes are smaller than at the upper end. This amounts to saying that the amplification is less at the lower end of the line than near the top. For the small signal chosen for illustration the variation in amplification over the cycle is negligibly small, but for a large signal it might cause serious distortion of the output. This has already been illustrated in a different manner by Fig. 4.9.

usually assume), then i_p and e_p will also be sinusoidal and we can label Fig. 4.13*b* with E_g, μE_g, I_p, and E_p, representing the effective values of the sine waves.

Although the equivalent circuit of Fig. 4.13 conveniently shows the relationships among the various alternating components, it may at first be disturbing to find that the voltages E_c and E_{bb} do not appear on the diagram. However, they do play a part because they determine the operating point, which in turn fixes the values of μ and r_p to be used. Thus the fixed quantities play the role of a stage setting on which takes place the action of interest.

4.7 Amplification with Resistive Load

Figure 4.14 shows the equivalent circuit for a triode with a resistive load. This is a simple series circuit for which we can

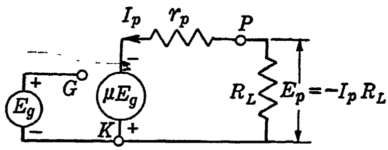


FIG. 4.14. Equivalent circuit for a triode amplifier with resistive load.

write from inspection

$$\mu E_g = I_p R_L + I_p r_p$$

$$I_p = \frac{\mu E_g}{r_p + R_L}$$

As discussed in connection with Eq. (4.11)

$$E_p = -I_p R_L = \frac{-\mu E_g R_L}{r_p + R_L}$$

The amplification of the circuit is the ratio between E_p and E_g. Therefore

$$A = \frac{E_p}{E_g} = \frac{-\mu R_L}{r_p + R_L} \tag{4.16}$$

This expression immediately shows that the amplification A approaches the amplification factor μ when the load resistance becomes large compared with the plate resistance. From a practical standpoint, however, it is not necessary to make the load resistance more than ten times the plate resistance in order to

realize 90 percent of the ultimate gain. Equation (4.16) also indicates that a high amplification factor should be associated with a low plate resistance to obtain high amplification with a reasonable size of load resistance. In other words the ratio μ/r_p should be large. Since μ/r_p is the transconductance, a high transconductance is desirable for a good amplifier tube.

It is interesting to use Eq. (4.16) to check the results obtained graphically from the load line. At the operating point the values of μ and r_p read from the curves are 20 and 10,000 ohms. The load line was drawn for an R_L of 30,000 ohms. Placing these values in the equation,

$$A = -\frac{(20)(30{,}000)}{(10{,}000 + 30{,}000)} = -15$$

This checks exactly the graphical analysis. The minus sign means that the output voltage is 180 degrees out of phase with the input, as shown by Fig. 4.10.

4.8 Amplification with an Impedance Load

Analyzing the behavior of an amplifier with a load impedance containing reactance is practically impossible graphically, but with the aid of the equivalent circuit it reduces to a simple series circuit problem. Figure 4.15 shows the equivalent circuit for a

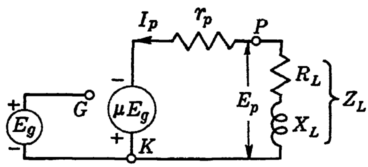


FIG. 4.15. Equivalent circuit for a triode amplifier with an inductive impedance load.

triode with an inductive impedance load. Treating this circuit exactly as before but substituting $\mathbf{Z}_L$ for R_L and remembering that the currents, voltages, and impedances must be treated as complex quantities, we obtain for the amplification

$$\mathbf{A} = -\frac{\mu \mathbf{Z}_L}{(r_p + \mathbf{Z}_L)} \qquad (4.17)$$

The bold-faced type indicates complex or vector quantities. Here again the minus sign represents a phase reversal of 180 degrees,

but the fraction $Z_L/(r_p + Z_L)$ is not a pure number and contains a phase angle. This means that the output voltage will be shifted from the normal 180 degrees by an amount depending on the reactance in the circuit.

The vector diagram of Fig. 4.16 illustrates this point. Voltages E_g and μE_g are used as the reference. In the plate circuit, the

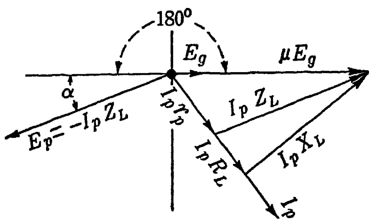


FIG. 4.16. Vector diagram for the amplifier of Fig. 4.15 with inductive load impedance.

current I_p lags the applied voltage μE_g because of the inductive reactance present. Voltage drops $I_p r_p$ and $I_p R_L$ are in phase with the current, as they must always be, but the IX drop in X_L leads the current by 90 degrees. Furthermore the sum of the three voltage drops must equal the applied voltage μE_g. This is shown on the diagram. Also shown is the drop $I_p Z_L$ obtained by adding the drops across R_L and X_L. Reversing the vector $I_p Z_L$ gives E_p which now is not 180 degrees out of phase with E_g. When this is the case, the amplifier is said to produce a phase shift of α degrees from the normal 180 degrees. Whether or not this phase shift is desirable depends entirely upon the particular circumstances. A circuit amplifying complex wave forms containing many different frequency components might produce a seriously distorted wave by disturbing the phase relations of the amplified components.

To illustrate a quantitative computation of the amplification with a reactive load, let us assume a circuit with a 6J5 triode and a load of 25,000 ohms inductive reactance and 5,000 ohms resistance. The impedance Z_L is then $5{,}000 + j25{,}000$. Placing this in Eq. (4.17),

$$A = -\frac{(20)(5{,}000 + j25{,}000)}{(10{,}000 + 5{,}000 + j25{,}000)} = -\frac{(20)(5 + j25)}{(15 + j25)}$$

Multiplying top and bottom by the conjugate to clear the fraction,

$$A = -\frac{(20)(5+j25)(15-j25)}{(15+j25)(15-j25)} = -\frac{(20)(700+j250)}{850}$$

This result was obtained by carrying out the multiplication and remembering that j^2 is minus one. Completing the computation,

$$A = -16.5 - j5.88 = 17.5\underline{/199.6°}$$

This final form is obtained with the aid of Fig. 4.17, illustrating the change from rectangular into polar form. For this particular

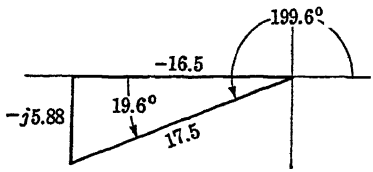


FIG. 4.17. Illustration of the change from rectangular into polar form.

amplifier the phase shift α is 19.6 degrees leading the normal 180 degrees.

4.9 Interelectrode Capacitance

Any two conductors near one another possess the property of electrical capacitance and, unfortunately, the elements of a triode are no exception. This is illustrated by Fig. 4.18, showing the three capacitances that exist between the cathode, grid, and plate. Capacitor C_{gk} between the cathode and grid appears as a reactance to the signal source; its effect will be discussed in connection with practical amplifier circuits. The output capacitance C_{pk} is effectively in parallel with the load impedance, and it can be considered as part of Z_L. The presence of the grid-plate capacitance, however, may be very annoying because C_{gp} provides an a-c path for the transfer of energy from the plate circuit to the input.

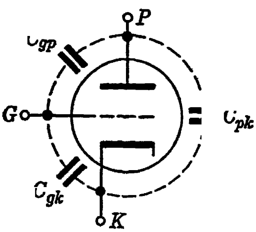


FIG. 4.18. Triode interelectrode capacitances.

The three capacitances are small, usually in the order of 5 micromicrofarads for the tube alone but somewhat larger when the tube socket and wiring are taken into account. At audio fre-

quencies the reactance is high; for 1,000 cycles per second the reactance of 5 micromicrofarads is over 30 megohms, but at a radio frequency of 1 megacycle (the middle of the broadcast band) the reactance is only 30,000 ohms. Thus at a moderate radio frequency the reactance of C_{gp} is comparable to the load resistance of the amplifier. Under such circumstances the capacitance provides a relatively good path for the transfer of amplified energy from the plate circuit back to the grid again. This usually results in a condition called oscillation, in which the amplifier provides its own input signal, and an output is obtained even when no input is applied. Since an amplifier is supposed to produce an output that faithfully follows the input, oscillation is undesirable and a triode cannot be used at high frequencies without taking special precautions to eliminate feedback.

4.10 The Tetrode

The tetrode (four-element tube) represents a first attempt to eliminate the grid-plate capacitance of the triode. As indicated by Fig. 4.19, this is accomplished by placing a second grid of wires between the first grid (called the control grid) and the anode. The second grid G_2 is called the screen grid because it intercepts the electrostatic field of the plate and reduces the capacitance between plate and G_1 to a very low value. Of course capacitance exists between G_2 and the plate, but by connecting G_2 to the cathode this capacitance is placed in parallel with the load impedance, where it does no particular harm. Unfortunately, connecting G_2 to the cathode would also reduce the plate current to zero because the screen grid in shielding the plate from the control grid also shields it from the cathode so that no electrons are attracted to the anode. This difficulty can be overcome by introducing a battery at point X of Fig. 4.19 to maintain G_2 at a constant positive potential with respect to the cathode.

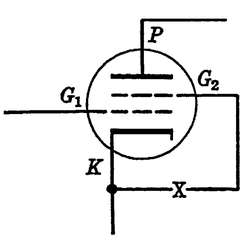


FIG. 4.19. Conventional symbol for a tetrode.

Let us now investigate the flight of an electron from the cathode to the anode. With G_2 positive, the electric fields between the cathode and G_2 are similar to those in a triode and a portion of the emitted electrons manage to pass the control grid and accelerate

toward the screen. Of these electrons, a fraction are captured by the screen itself, but those whose paths take them between the screen wires pass it and continue toward the plate. The situation can best be described with the aid of Fig. 4.20, showing a plot of potential as a function of distance from the cathode. The heavy line shows the voltage for a path between the grid wires (the path in which we are interested), and the dotted line shows the potential curve for a path intersecting the grid wires. Three cases are of interest as follows:

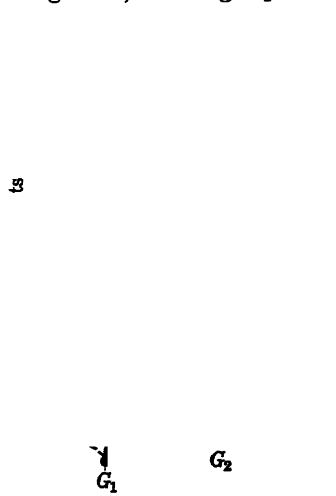


FIG. 4.20. Potential distribution in a tetrode as a function of the distance from the cathode. The heavy line represents an electron path passing between the grid wires. The dotted line shows the potential along a path intersecting a grid wire.

Plate Voltage Higher Than Screen Voltage. Under this condition the electrons continue to accelerate toward the anode, and all those passing the screen reach the plate. Upon striking the plate the primary electrons may cause the emission of secondaries, but these return to the plate and contribute nothing to the operation of the tube.

Plate Voltage about Equal to Screen Voltage. In this case the electrons passing the screen continue to coast on to the plate with little velocity change. Secondary electrons driven out of the anode may or may not leave the plate and reach the screen, depending on whether their kinetic energy is sufficient to carry them under the slight dip in curve *b* of Fig. 4.20. This dip is caused by the space charge.

Plate Voltage below Screen Voltage. When the plate voltage is considerably lower than the screen voltage, the electrons are *decelerated* after passing the screen but reach the plate because of the kinetic energy acquired in traveling from the cathode to the plane of the screen. Any secondary electrons produced find an electric field urging them away from the plate and toward the screen. Thus they continue toward the screen, and the net plate current equals the current due to the primary electrons *minus* the secondary

electron flow. If more than one secondary electron is produced per primary electron, the plate current may actually be negative.

The effect of secondary emission on the tetrode plate characteristic curves is shown by Fig. 4.21. The dotted line shows the *primary* current reaching the plate. This curve rises rapidly at low plate voltages and levels off when all the electrons passing the screen reach the anode. Since the screen cannot completely intercept the electrostatic field of the anode, the plate potential does have a slight effect on the current and the curve is not perfectly

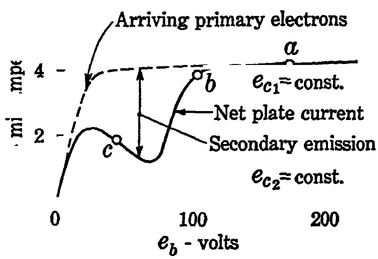


FIG. 4.21. Analysis of a tetrode plate-current curve. The curve is taken with constant screen and control-grid voltages.

level. At very low plate voltages primary electrons strike the plate so gently that secondaries are not formed. Consequently, the *net* plate current approaches the dotted line near the origin. For higher plate voltages the kinetic energy of the primary electrons is greater, secondary electrons are formed, and the net plate-current curve departs from the upper curve. In the neighborhood of *c*, where the primary current curve is flat, the rate of increase of secondaries makes the net current drop off until a point is reached where the plate potential begins to approach the screen voltage. In this region (below *b*) the space-charge dip between the screen and the plate discourages the weaker secondary electrons, and the net current again approaches the upper curve. Well above *b* all the secondaries return to the plate, and the curves again coincide.

The tetrode plate-current curve of Fig. 4.21 is not without its interesting and useful features, but for the purpose of amplification the portion affected by secondary emission is undesirable. Of course this part of the curve can be avoided by using a sufficiently high plate-supply voltage E_{bb} to make the load line fall to the right of the critical region, but this amounts to wasting the first

one hundred volts or so of d-c supply—hardly an economical procedure. Consequently, the tetrode was soon abandoned in favor of the pentode.

4.11 The Pentode

A pentode consists of a tetrode with a third grid G_3 added between the screen grid and the plate as suggested by the conventional diagram of Fig. 4.22. This grid, called the suppressor, is designed to suppress the effects of secondary emission by repelling the secondary electrons back to the plate instead of allowing them to reach the screen. Figure 4.23 illustrates the role of the suppressor in accomplishing this result. The diagram shows that the suppressor lowers the potential distribution curve in front of the plate and forces the secondary electrons to return. Ordinarily the suppressor is connected to the cathode, but in some cases a small positive or negative potential may be applied.

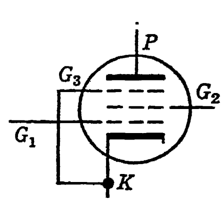


FIG. 4.22. Conventional symbol for a pentode.

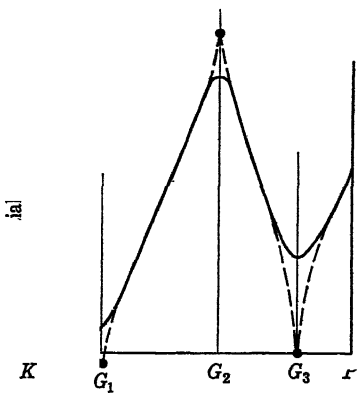


FIG. 4.23. Potential distribution curve for a pentode.

Figure 4.24 shows the resulting family of plate characteristics for a type 6SJ7 pentode designed for voltage amplification. No trace of secondary emission dip can be found, and each curve first rises rapidly and then levels off rather abruptly. Not all pentode curves break as sharply as these; those designed for higher plate currents and greater power are usually more rounded.

Pentode Coefficients. The transconductance of a pentode is approximately the same as that of the corresponding triode. For

the 6SJ7 pentode the transconductance at a plate current of 4 milliamperes is 1,800 micromhos as compared with 2,000 for the 6J5 triode. The plate resistance of a pentode is invariably high because the plate voltage has little effect on the plate current in the normal operating range. The measured slope of the flat part of the 6SJ7 curves shows the plate resistance to be in the order of several megohms, but the value varies considerably with the position on the chart. While high plate resistance is sometimes undesirable, it is a direct consequence of the introduction of shielding between the plate and the control grid.

It is not possible to obtain the amplification factor directly from the curves but, with the help of the relation $\mu = g_m r_p$, it can be estimated. Putting in the values of transconductance and plate resistance we find that the amplification factor is very high—in the order of 4,000. This is a tremendous advantage because the amplification of a circuit cannot exceed the amplification factor. With a pentode, the amplification factor is so high that a circuit realizing only a small fraction of the ultimate may have a gain of over a hundred.

Pentode Amplification. The load line drawn on Fig. 4.24 corresponds exactly to that drawn on Fig. 4.5 for the 6J5 triode. For the pentode a 2-volt swing from -1 to -3 on the grid produces a plate-voltage change of 116 volts. This represents an amplification of 58 for the pentode as compared with 15 for the triode under similar conditions of operation. Thus the effort spent in reducing the undesirable grid-plate capacitance has also improved the tube from the standpoint of amplification. This represents one of those rather rare circumstances in which the removal of one defect results in the unexpected improvement of another characteristic.

The amplification of an amplifier is given by the expression

$$A = -\frac{\mu Z_L}{r_p + Z_L} \tag{4.18}$$

However, the plate resistance of a pentode is much larger than any practical load impedance or resistor capable of passing the necessary plate current with a reasonable voltage drop. Therefore, to a close approximation, the term Z_L in the denominator of the above expression can be neglected with the result

$$A = -\frac{\mu Z_L}{r_p} = -g_m Z_L \tag{4.19}$$

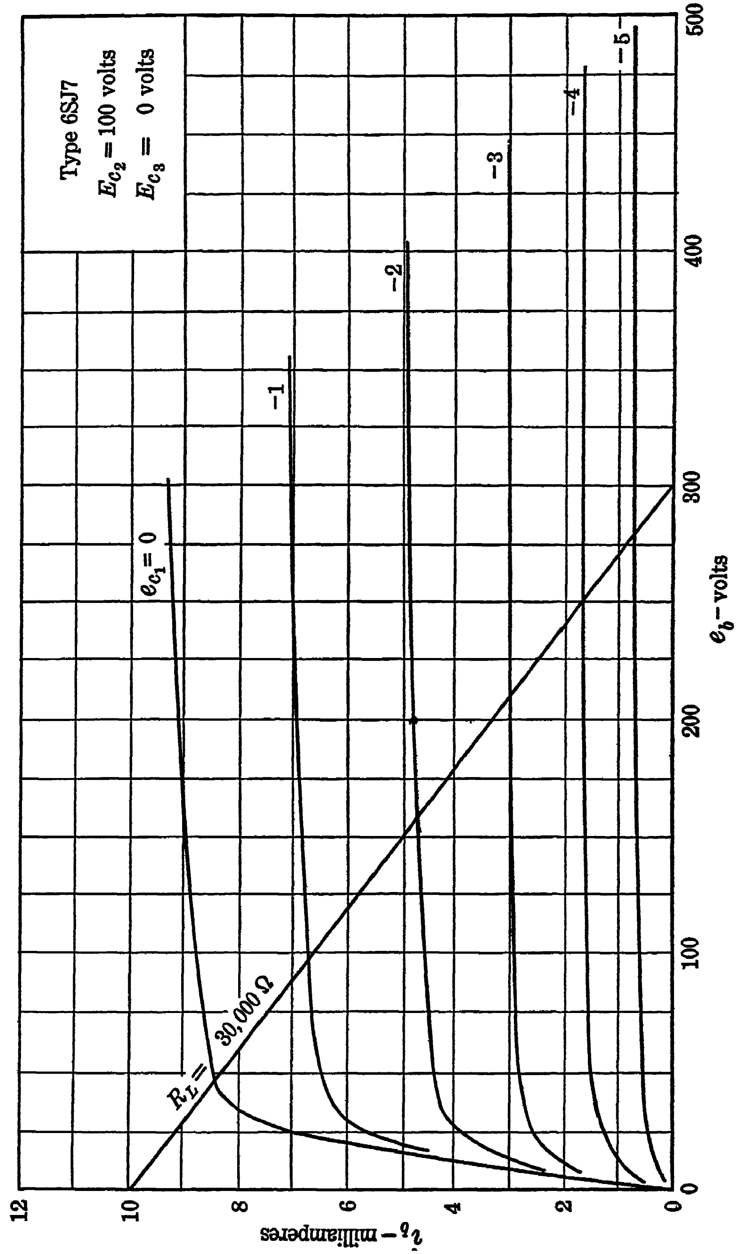


FIG. 4.24. A family of plate characteristics for a type 6SJ7 pentode.

This means that, within the practical range of load impedance, the amplification is proportional to the load impedance rather than limited by a relatively low amplification factor. Figure 4.25 illustrates this point. Equation (4.19) also shows that the transconductance is the only factor of interest for a pentode as long as the plate resistance is very high.

Investigation of the load-line intercepts on the characteristic curves of Fig. 4.24 shows them to be less uniform than for a triode. Therefore the distortion of the output wave is worse if

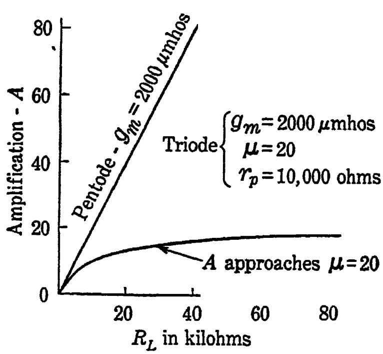


FIG. 4.25. Comparison of the amplifications obtained with a triode and a pentode of equal transconductance.

pentodes are used in place of triodes. Although this is generally true, pentodes have so many other outstanding advantages that they have practically displaced triodes for many applications. Fortunately there are other ways of reducing the nonlinear distortion produced by an amplifier.

4.12 The Beam-power Tetrode

The beam-power tube is essentially a tetrode of improved design to eliminate the effects of secondary emission by taking advantage of the electron space charge existing between the screen and the plate. In conventional tubes the electrons start at a small cylindrical cathode and spread out more or less radially as they flow to the anode. This makes the electron density near the plate small, and the space charge is too weak to be of any particular help in repelling secondary electrons back to the plate. By arranging the electrodes in the fashion shown by Fig. 4.26, it is possible to obtain a nearly parallel flow and maintain a high-density beam throughout the path. Hence the term "beam" tube. The elec-

trons start from a flattened cathode and are prevented from spreading radially by beam-forming plates located on each side

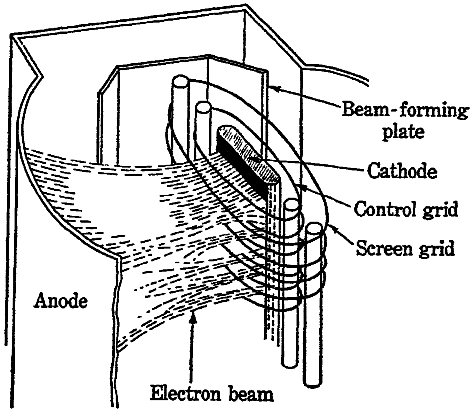


FIG. 4.26. Cutaway view of a beam-power tetrode to show the formation of the electron beam.

of the path and connected to the cathode. An additional improvement is obtained by winding both grids to the same pitch and carefully aligning the screen grid so that it falls in the "shadow"

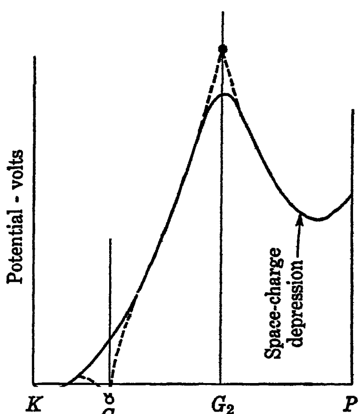


FIG. 4.27. Potential distribution in a beam-power tetrode.

of the negative control grid. This reduces the screen current and improves the tube efficiency.

Figure 4.27 shows the potential distribution in a beam tube with

a plate voltage below the screen voltage. Electrons passing the screen are decelerated as they approach the plate and form a dense space charge. This produces a potential depression and provides a barrier to the flow of secondary electrons from plate to screen. Imagine a lone secondary electron in such a position confronted by a crowd of repulsive primary electrons rushing toward it. At low plate voltages the primary electrons are slowed down even more and produce a still lower potential minimum. Thus secondary emission is effectively suppressed without a third grid.

The characteristic curves for a small beam-power tube are given by Fig. 4.28. These curves show good design because the plate current rises rapidly until the knee of the curve is reached. This permits a maximum plate-voltage swing along the load line without driving the control grid positive. Since this provides the utmost in alternating output for a given d-c input, a beam-power amplifier is relatively efficient as well as providing more amplification than does a triode. The lower curves show an interesting and typical tetrode secondary-emission dip near the left-hand end. At low currents the small space charge cannot depress the potential enough to return the secondaries to the anode. Fortunately, a normal load line falls well above this region.

The beam-tetrode design is used only for power-amplifier tubes where currents are high and efficiency is important. Most voltage amplifiers operate at such small currents that efficiency is secondary, and pentodes do a satisfactory job. In beam tubes the role of the screen as a shielding electrode is often secondary; its main function is to accelerate the electrons to reach the plate even with low plate voltages.

PROBLEMS

4.1 From the characteristic curves of Fig. 4.5, graphically determine the plate resistance, transconductance, and amplification factor of a 6J5 triode for a number of plate currents between zero and 12 ma. Determine these for plate voltages in the vicinity of 200 volts, and plot curves showing the three coefficients as a function of the plate current.

4.2 On the chart of Fig. 4.5, draw a load line for the amplifier of Fig. 4.8 with $E_{bb} = 300$ volts, $E_c = -8$ volts, and $R_L = 20{,}000$ ohms. Determine I_b, E_b, and the amplification in the vicinity of the operating point.

4.3 From the load line of Prob. 4.2, draw the dynamic characteristic and develop a plate-current wave similar to that of Fig. 4.9. Assume a sinusoidal grid voltage with an 8-volt peak.

4.4 A type 6J5 in the circuit of Fig. 4.8 operates with an actual plate voltage

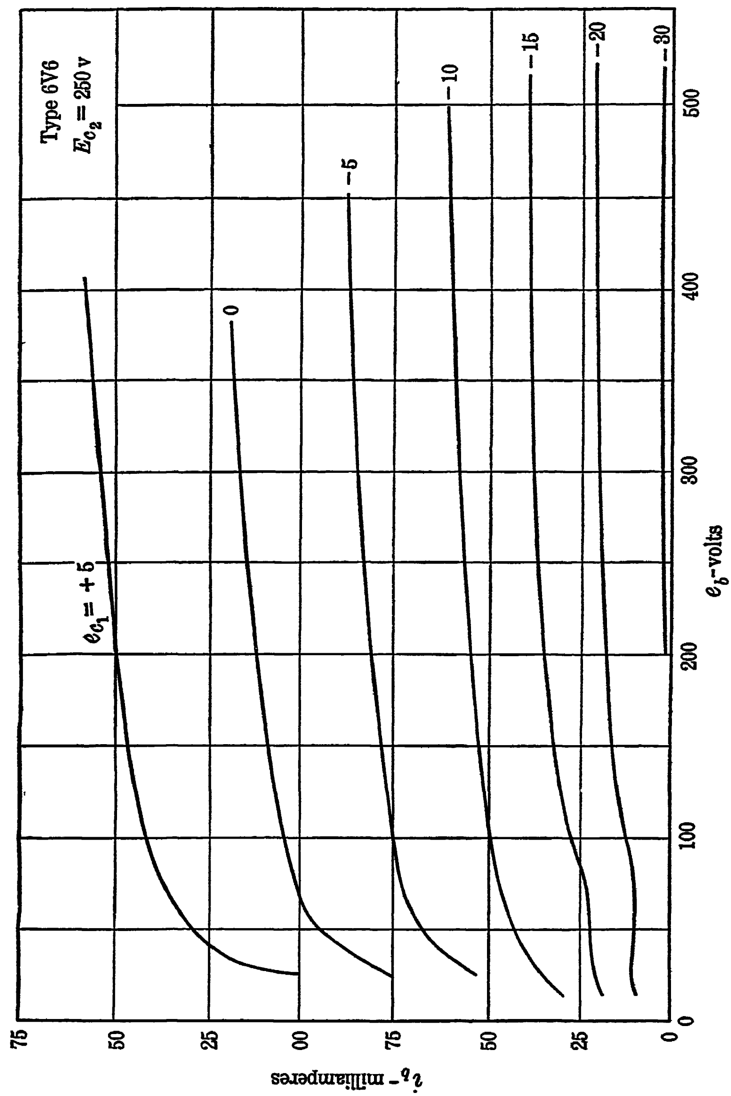


FIG. 4.28. Plate characteristics for a type 6V6 beam-power tetrode.

of 150 volts (E_b), a bias of -5 volts, and a load resistance of 50,000 ohms. Draw the load line for the circuit, and from it determine the required power-supply voltage E_{bb} and the amplification.

4.5 A high-mu triode operates into an inductive load impedance $Z_L = 1{,}000 + j25{,}000$ ohms. The tube coefficients are $\mu = 100$, $r_p = 80{,}000$ ohms. Compute the amplification, and determine the relative phase angle between E_p and E_g.

4.6 A triode with a μ of 20 and an r_p of 10,000 ohms operates with a load impedance of 10,000 ohms resistance in parallel with a capacitance of 100 $\mu\mu$f. Compute (*a*) the frequency at which the capacitive reactance equals the 10,000-ohm resistance, and (*b*) the amplification at this particular frequency.

4.7 Compute the amplification of the circuit of Prob. 4.6 at one-tenth of the frequency at which $X_c = R$.

4.8 From the curves of Fig. 4.24, estimate the value of the plate resistance for a type 6SJ7 at the point $E_b = 200$ volts, $E_c = -2$. Also determine the transconductance at this point, and compute an approximate value of the amplification factor.

4.9 Draw a 40,000-ohm load line through the operating point of Prob. 4.8, and determine the amplification (*a*) graphically, (*b*) from the exact equation using the values of Prob. 4.8, and (*c*) from the approximate equation for a pentode amplifier.

CHAPTER 5

GAS–FILLED TUBES

ELECTRICAL discharge in gases is a complex phenomenon, and many books have been written about the subject. Even today, with the fundamental processes well understood, a complete quantitative analysis is not always possible.

For the purpose of investigation, let us imagine a glass envelope containing two cold metal electrodes and a gas at a pressure considerably below atmospheric. Figure 5.1 shows this tube connected in a test circuit with suitable meters and a series resistor to protect them when the gas breaks down. The heavy dot inside the envelope shows the conventional representation for gas.

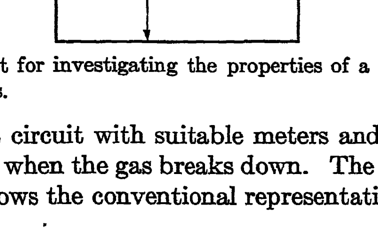


FIG. 5.1. A circuit for investigating the properties of a gas-filled tube with two cold electrodes.

5.1 The Townsend Discharge

By varying the plate-supply voltage from zero upward and taking simultaneous readings of current and voltage, we can obtain the data for the curve of Fig. 5.2. This curve shows that the plate current first increases up to point a and then levels off for a short distance up to point b. Under normal circumstances this current is extremely small and, from a practical standpoint, the tube remains an excellent insulator, but from the physical point of view the current is extremely interesting. This tiny current results from

ionization of the gas by photons and high-energy particles, and from photoelectric emission at the cathode. Many experiments substantiate this fact. Exposing the tube to ionizing X rays or radioactive material immediately increases the current without changing the general shape of the curve. Ultraviolet light shining on the cathode causes a similar increase by photoemission. Surrounding the tube with a shield of lead or thick concrete decreases the current but does not reduce it to zero. High-energy cosmic rays penetrate such a barrier to ionize the gas.

The shape of the curve is now easily explained. A voltage above

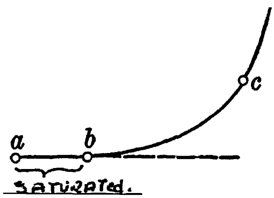


FIG. 5.2. The Townsend discharge curve obtained with the circuit of Fig. 5.1.

that of point *a* saturates the tube and draws all the electrons and positive ions to the electrodes as fast as they are formed. Increasing the potential to *b* has no effect on the ion supply; the individual particles acquire higher velocities, but the number arriving per second at the electrodes remains the same. For voltages higher than *b*, however, the electrons acquire enough energy to produce additional ions through collisions with the gas molecules. The bulky positive ions do not contribute to this because they continually collide with gas atoms of equal mass and lose so much of their kinetic energy at each collision that they gain insufficient velocity to cause ionization. The tiny electrons, on the other hand, travel relatively large distances between collisions, and being so light compared with the gas atoms, they lose little energy from the impacts. They thus continue to gain energy from the electric field until they have sufficient energy to produce ionization.

At point *c* on the curve, each electron produces, on the average, *two* additional ionizing collisions and increases the current to *three times* the initial value. An analysis of this process indicates that

the current should increase exponentially with applied voltage. Electron "avalanches" are formed in which the initiating electron liberates an additional electron; the two then speed off to repeat the process, making four electrons, and so on. From this picture it is not surprising that the current curve rises very steeply, but there is no reason to expect that the current should eventually

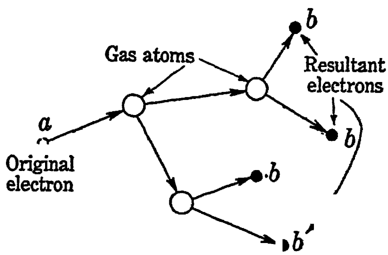


FIG. 5.3. The formation of an electron avalanche by successive ionization in a gas.

rise without limit because each electron can produce only a finite number of additional electrons. Experimentally, however, the current does reach a point where the slope becomes vertical.

This was investigated by Dr. J. S. Townsend, for whom the discharge is named, as early as 1901. Townsend reasoned that the electron avalanches alone could not produce such a curve and that the positive ions traveling toward the cathode must somehow produce additional electrons in a favorable position near the cathode to start fresh avalanches. He assumed that the ions produced ionization by collision with gas atoms, but it has since been shown that the additional electrons are liberated by secondary emission when the ions strike the cathode. In a typical discharge only one ion in several thousand produces a secondary electron.

At the critical voltage where the curve rises without limit, each original electron produces sufficient ions so that they, in turn, can produce one new electron at the cathode. At this point the discharge becomes *self-sustaining* because it no longer depends upon the initial source of ionization. The voltage at this point is called the breakdown voltage or the sparking potential. Below this point the discharge is *non-self-maintaining*.

A study of the Townsend discharge is important because it illustrates the mechanism by which electrical breakdown occurs in a gas. The discharge itself is important in the operation of gas-

filled phototubes, which take advantage of the current amplification provided by the electron avalanches.

5.2 The Glow Discharge

The series resistance in the circuit of Fig. 5.1 prevents the current from increasing without limit when the sparking potential is reached. Instead, it is found that, as this potential is approached, the current suddenly takes a jump, the voltage across the tube drops, and a glow appears. This is called a *glow discharge*. Further increases in supply voltage increase the tube current but may actually decrease E_b. Consequently, it is more convenient to think of the current as the independent variable and the voltage as the dependent function. For this reason, Fig. 5.4 has been

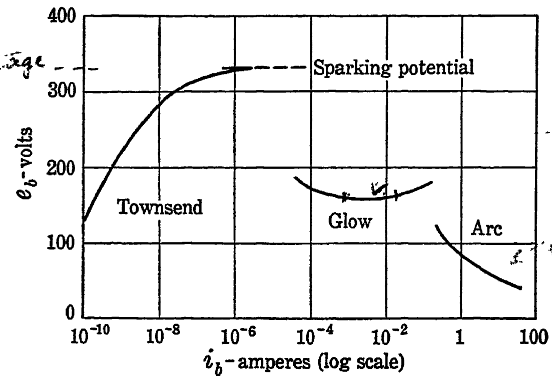


FIG. 5.4. Voltage-current characteristic curves for the three types of gaseous discharge.

plotted with the current as the abscissa. The logarithmic scale shows the wide current range covered by the three types of discharge illustrated.

The glow discharge is a *self-maintaining* discharge characterized by a relatively constant voltage drop over a considerable range of current. The mechanism is considered to be the same as for the Townsend discharge as it approaches the self-sustaining point; positive ions striking the cathode produce secondary electrons which start electron avalanches, thus producing more positive ions, etc. The important difference between the Townsend and glow discharge is the current carried. In the glow this current is large enough to produce effective space charges that distort the field

in the space between the plates. This accounts for the difference in voltage between the two discharges despite the similarity of the mechanisms involved. The region between the two curves is an unstable one, and in most cases it cannot be observed experimentally.

The flat portion of the glow-discharge curve is of practical importance because it provides a device that will maintain a fairly constant voltage drop despite current variations through it.

5.3 The Arc Discharge

The positive ions striking the cathode in a glow discharge not only liberate secondary electrons, but they also heat the surface. Normal glow currents produce only moderate warming of the cathode, but if the current is increased sufficiently, a point is finally reached at which the cathode temperature becomes high enough to produce thermionic emission. This completely changes the character of the discharge, and again there is a sudden jump to a new form of conduction called an arc.

The arc is characterized by a relatively low voltage drop and a relatively high current, and it differs from the glow and Townsend discharges by the presence of a copious electron supply at the cathode.

Thermionic Arcs. In an arc, heating of the cathode by positive-ion bombardment produces thermionic emission far in excess of the secondary-emission capabilities of the ions. In the glow, many ions are required to produce a single secondary electron, but when thermionic emission occurs, many electrons are liberated per ion. This, in turn, means that each electron in traveling toward the anode needs to produce only a fraction of an ion, on the average, instead of making many ions through electron avalanches. For this reason, the arc can exist at the rather low voltage drops sufficient to accelerate the ions just enough to supply the required cathode heating. In fact, thermionic emission increases so rapidly with the temperature that the ratio of electron emission to heating energy decreases at larger currents. Thus, for an increased current flow, each positive ion needs to contribute less energy than before (although the total will be more), and the voltage drop across the arc decreases. Quantitative computations for refractory cathodes, such as tungsten, provide surprisingly good verification of the experimental voltage-current curves.

The discharge obtained when the cathode is heated by some separate means is also classified as an arc. This heating has the advantage of eliminating the progression of the discharge through the Townsend and glow stages, and it permits the arc to start when the voltage reaches the ionizing potential of the gas without having first to override the sparking potential. Since the energy for heating the cathode is not provided by positive-ion bombardment, it is to be expected that the voltage drop for this type of arc may be even lower than shown by Fig. 5.4. In fact, the voltage drop for this arc is approximately equal to the ionizing potential of the gas since the electrons need only to reach sufficiently high velocities to produce ionization. With mercury vapor as the gas, a tube can be constructed that is capable of passing large currents with a voltage drop of only 10 volts.

Cold-cathode Arcs. Arcs in which the cathode has a relatively low melting point apparently do not depend on thermionic emission. Mercury, for instance, is commonly employed for arc cathodes; yet it vaporizes at a temperature far below that necessary to produce thermionic emission. Furthermore, traveling arcs can be produced in which the cathode spot moves rapidly along the metal surface without time to heat it appreciably. Examination of a copper cathode after such an arc has passed may show no evidence of pitting or burning, yet the melting point of copper is below the emission temperature.

It is now generally believed that such arcs depend on some form of high-field emission caused by the heavy concentration of positive ions at the cathode surface. Theoretical computations indicate that the cathode layer of positive ions can produce the extreme field intensities required.

5.4 The Glow Diode—Voltage Regulator

One important application of the glow discharge is the glow diode, or voltage-regulator, tube. This tube takes advantage of the flat portion of the glow-voltage characteristic to provide a device with a voltage drop substantially independent of the current carried. In its commercial form, the tube consists of a cylindrical cathode surrounding a central anode rod. Following evacuation, gas at low pressure is introduced. The inert gases helium, neon, argon, etc., are employed because they ionize easily and they do not react with the electrodes. Even with the inert gases there is a

gradual adsorption by the electrodes, and the tube characteristics slowly change with time.

Figure 5.5 shows the characteristic curve for a commercial voltage-regulator glow tube designed for 105 volts. This voltage is maintained within a few percent over the operating range of 5 to

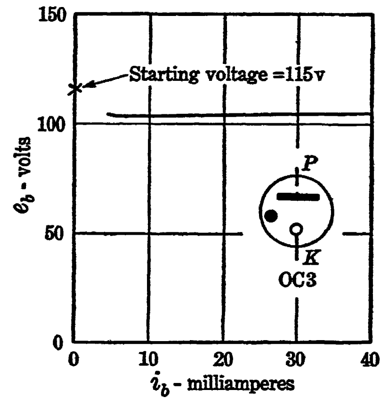


FIG. 5.5. Characteristic curve for a type OC3 voltage-regulator glow tube.

40 milliamperes. By suitable spacing of the concentric electrodes, the breakdown voltage (starting voltage) is made rather low and it is unnecessary to apply a high voltage to start the glow. Other tubes are available for voltages of 75, 90, and 150 volts with similar current ratings.

These glow tubes are used in voltage-regulator circuits designed to provide a constant output voltage regardless of load-current or input-voltage changes. Figure 5.6 shows such a circuit. Here,

$I_L + I_{tube} =$

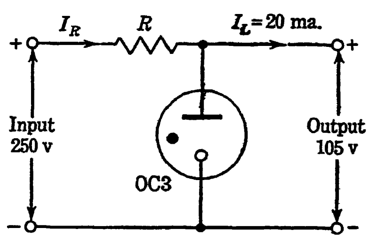


FIG. 5.6. Simple voltage regulator employing a glow diode.

the problem is to provide a constant 105-volt source of 20 milliamperes from the 250-volt output of a rectifier. This regulated

voltage might be used for the screen voltage of a number of pentodes in a circuit requiring especially good stability. The circuit design is based on the fact that the output voltage will remain approximately 105 volts as long as the glow-tube current stays within the range of 5 to 40 milliamperes. Choosing a glow-tube current of 20 milliamperes sets the value of I_R at 40 milliamperes. Resistor R must then be $(250 - 105)/0.04$, or 3,600 ohms. The nearest commercial size will do.

If, by chance, the load is removed, the output voltage still remains 105 volts and the current I_R remains the same. Therefore the glow-tube current must make up the difference by increasing to 40 milliamperes. Increasing the load current to more than 35 milliamperes drops the glow-tube current below 5 milliamperes, which is outside the operating range. Changes in input voltage also affect the glow-tube current without appreciably changing the output voltage, as long as the tube current stays between 5 and 40 milliamperes.

The voltage-regulator circuit is also an excellent filter for reducing the input-voltage ripple. If, for instance, a 100-volt change of input voltage produces only a 1-volt change in the output, the equivalent smoothing factor is 100.

5.5 The Thermionic Gas Diode

One application of the arc discharge is the thermionic gas diode used instead of the vacuum diode for efficient rectifier circuits. In its common form, the diode consists of an oxide-coated cathode and an anode, both of which are enclosed in an envelope containing a drop of mercury. Other inert gases may be used, but the mercury vapor has the advantage of low ionizing potential, and the liquid mercury provides a source of gas to replace any adsorbed by the electrodes. However, the excess mercury has the disadvantage of making the gas pressure and the tube characteristics dependent on the temperature.

Characteristic Curves. Figure 5.7 shows the characteristic curve of a mercury-vapor diode. For voltages up to point a, the tube acts much like a vacuum diode. The electrons have insufficient speed to make positive ions, and the current is limited by the negative space charge. At approximately the ionizing potential of the gas (10.4 volts for mercury) the electrons succeed in producing positive ions. Mercury ions are extremely heavy compared

with electrons, and they move slowly toward the cathode, remaining in the intervening space a relatively long time. For this reason, the ions are highly effective in counteracting the electron

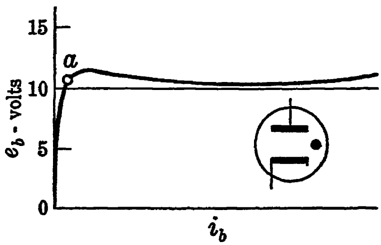


FIG. 5.7. Characteristic curve for a hot-cathode mercury-vapor diode.

space charge, and the current is no longer limited by space charge; instead, it is limited only by the external circuit resistance. This condition is represented by the relatively flat portion of the curve of Fig. 5.7.

The effect of the positive-ion space charge on the potential distribution in the tube is shown by Fig. 5.8. From the cathode to

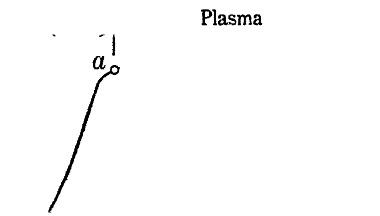


FIG. 5.8. Potential distribution in a thermionic gas diode.

point a the potential rises rapidly, as it must, to accelerate the electrons up to ionizing velocity. From point a onward, the electrons produce ionizations, and the remaining distance to the anode is filled with ions and electrons of approximately equal number. Thus this region has all the properties of a good conductor—zero net charge and free conduction electrons. This portion of the discharge is called the plasma, and the voltage drop throughout its length is very small.

If the anode is moved farther away from the cathode, the potential distribution in the vicinity of the cathode is not much affected; the plasma becomes longer, and the tube drop is about the same.

The net effect is the same as though the anode were very close to the cathode, with the plasma bridging the remaining distance. With more current flowing, the plasma moves even closer and maintains the same total voltage drop. Distance d of Fig. 5.8 is a few tenths of a millimeter for normal current densities.

Heat Shielding. This behavior immediately suggests the possibility of folding or roughening the cathode surface to increase the emission area without increasing the radiation loss. Since the plasma accurately follows the cathode contours, the tube drop is not particularly affected.

Figure 5.9 shows several heat-shielded cathodes. The directly

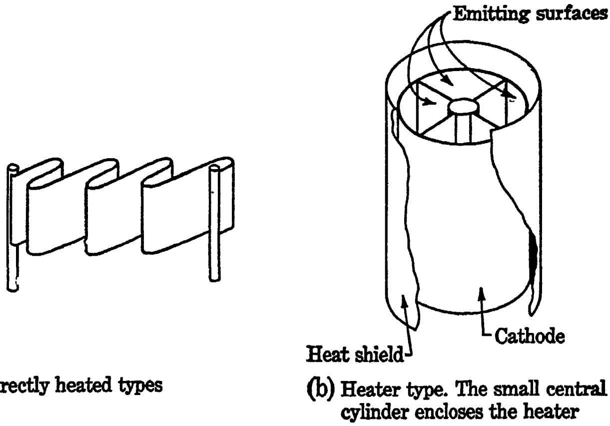


FIG. 5.9. Several types of heat-shielded cathodes. These improve the cathode efficiency by increasing the emitting surface without a corresponding increase in the radiating area.

heated types of Fig. 5.9*a* are commonly used in the smaller gas-filled tubes. The folded or coiled filament may either be solid or constructed of screen to increase the surface area further. Such an arrangement increases the emission per watt of heating power several times compared with conventional filaments. Even better is the heater type of Fig. 5.9*b*. The central heater heats the two inner cylinders and connecting vanes to emission temperature. This part of the structure is oxide-coated for maximum emission. Surrounding the emitter, one or more polished concentric heat shields reflect the radiant energy back to the coated cathode. Since the gas pressure is very low and convection plays little part, almost all the heat loss takes place from the ends of the structure.

The effort spent in shielding the cathode to reduce the required heating power also increases the thermal inertia of the system, with the result that normal heating power may require a long time to bring the cathode up to operating temperature. Many heater-type shielded cathodes require a 5-minute heating period. Directly heated cathodes are quicker; about 30 seconds is usually required.

Cathode Disintegration. It is perhaps surprising that oxide-coated cathodes can be used in gas-filled tubes because of the possibility of disintegration by positive-ion bombardment. Experimentally it has been found that appreciable cathode destruction can be avoided by keeping the tube drop below about 20 volts. Unfortunately, however, the tube drop depends on the temperature, especially for mercury-vapor tubes in which the gas pressure provided by the mercury drop is an exponential function of the condensed mercury temperature. Figure 5.10 shows that the tube

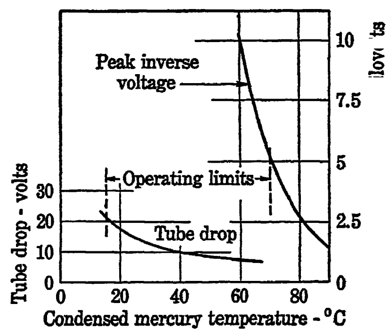


FIG. 5.10. Tube drop and peak-inverse-voltage curves for a mercury-vapor thermionic diode. These curves show that the condensed mercury temperature must remain within definite limits to avoid cathode disintegration and arc back.

drop increases at low temperatures and that the temperature of the mercury must be kept above a limiting value to avoid cathode disintegration. At low ambient temperatures merely heating the cathode for the specified time does not assure proper mercury temperature; the tube as a whole must be warm. Inert-gas-filled tubes can operate over wider extremes of temperature because the pressure of a free gas varies less than the pressure of a vapor in contact with the liquid phase.

Under normal circumstances the voltage drop across the tube is not much affected by the current through it as long as the current drawn is less than the total cathode emission. When saturation is reached, however, the tube drop rises and quickly exceeds the cathode disintegration potential. This limits the peak currents that can be drawn from a thermionic gas tube without damage and makes it imperative to bring the cathode to full operating temperature before plate voltage is applied. If plate potential is applied earlier, the plate current may attempt to exceed the emission of the partially heated cathode and produce excessive tube drop. For this reason, these tubes are usually protected by a time-delay relay to prevent application of the plate potential until the cathodes have heated.

Arc Back. The mercury-vapor diode suffers from another disadvantage not common to the vacuum diode. In rectifiers and control circuits, the tube spends part of the time with a negative plate voltage, and it is important that no conduction take place during this period. With reversed voltage the presence of the gas makes it possible for a discharge to take place when the potential exceeds the sparking potential. With low circuit resistance, this discharge immediately passes to the arc stage and the tube becomes an excellent conductor. This is called arc back. Arc back in the rectifier circuit of Fig. 3.5 would short-circuit the supply transformer and operate any fuses or circuit breakers protecting the system.

Fortunately, the electrode configuration and gas pressure can be adjusted to provide peak inverse voltages in excess of 20,000 volts. However, with mercury-vapor tubes this voltage is affected by the condensed mercury temperature, as shown by Fig. 5.10. High temperature, and therefore high pressure, severely limits the inverse voltage capabilities of the tube and restricts the operating range, as shown on the chart. Since the gas pressure is regulated by the temperature of the condensed mercury at the bottom of the tube, it is only this region that must be cooled. Blowers are usually arranged so as to keep this portion of the tube coolest.

Field of Use. Thermionic gas diodes are used in practically all rectifying equipment handling sufficient power to make efficiency important and large enough to make the required time-delay and temperature-control equipment economically worth while. Small rectifiers usually employ the more rugged vacuum diodes because

serviceability is of more importance than efficiency. Very high-voltage equipment must also employ vacuum tubes because of the peak-inverse-voltage limitations of the gas diodes. Industrial equipment of moderate voltage and high output power may employ mercury-pool arc tubes to obtain certain advantages of ruggedness discussed in the following sections.

5.6 The Thermionic Gas Triode—The Thyratron

The term thyratron describes a thermionic gas tube having one or more grids for the purpose of controlling the flow of plate current. Although the control electrodes are called grids, they are often made in the form of solid cylinders and baffles similar to the construction shown by Fig. 5.11. In a vacuum tube this would produce very high tube drop, but in a gas tube the plasma extends to the cathode and the tube drop is low.

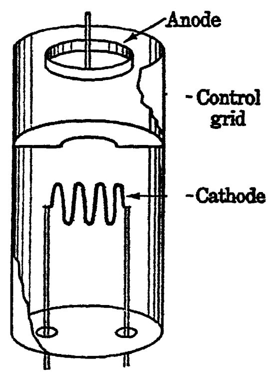


FIG. 5.11. Simplified sketch showing the construction of a thyratron or thermionic gas triode.

A thyratron has all the limitations of a gas diode with respect to operating temperature, cathode heating, etc., and it may depend on mercury vapor as a source of gas, or it may be filled with an inert gas such as argon to stabilize the characteristics with respect to temperature. Likewise it has all the efficiency advantages of a gas diode with the added feature of grid control.

Control Characteristics. The behavior of a gas triode is entirely different from that of a vacuum triode. It operates as an electronic switch that can be grid controlled with the expenditure of little power. This *control characteristic* is expressed as a curve, shown by Fig. 5.12. The heavy curve represents an average tube, and the dotted lines show the normal deviations from average. To understand this curve, let us imagine that the plate voltage of the tube is held constant at 200 volts, as indicated by the horizontal arrow on the diagram. The grid voltage is initially set at, say, −10 volts to produce complete current cutoff. Under this condition no electrons can accelerate toward the plate to produce ionization. Now let us gradually reduce the negative grid voltage until point a is approached. Up to point a, cutoff is maintained and the tube cannot conduct, but as soon as point a is passed, a tiny current

starts to flow to the plate. These electrons immediately produce positive ions which travel toward the negative grid and raise the potential near it. This permits more electrons to flow, more ions to be made, and the tube suddenly becomes fully conducting. The

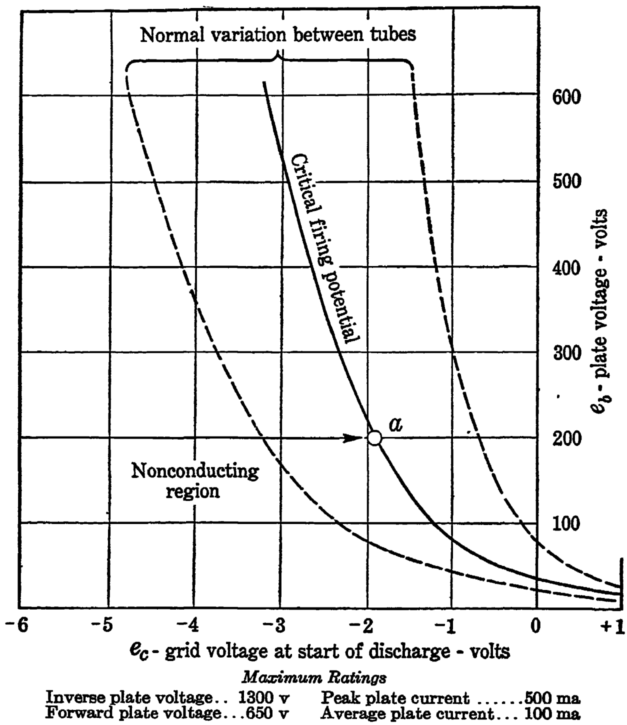


FIG. 5.12. Control characteristic and ratings of a type 2050 thyratron. This is a small thyratron similar in size and appearance to an ordinary radio receiving tube.

plasma now extends close to the cathode, and the grid potential has practically no effect on the tube drop or plate current.

Figure 5.12 shows that a more negative grid voltage will prevent firing with higher applied plate voltages. The lower left-hand region of this chart represents combinations of plate and grid voltages that can prevent conduction, and the remaining region shows combinations for which the tube conducts.

Thyratrons can be divided into two classes, depending on the position of the control characteristic. The type 2050 of Fig. 5.12

is called a negative-control type because the critical grid-voltage values are all negative for the higher anode voltages. Such a tube always fires with no voltage on the grid. Figure 5.13 also

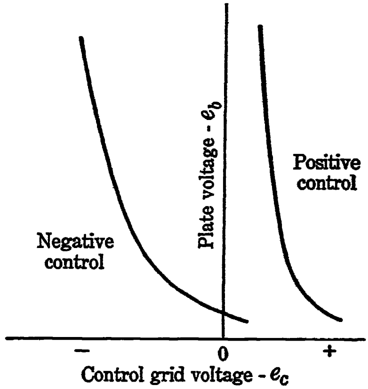


FIG. 5.13. A comparison of positive- and negative-thyratron-control characteristics.

shows a positive-control characteristic. A positive-control tube does not conduct with zero voltage on the grid, and it is necessary to apply a positive potential to initiate current flow.

Deionization. Once the tube has fired, it cannot ordinarily be stopped by making the grid more negative than the critical firing potential. Instead, positive ions flow to the negative grid and form a dense positive sheath that practically cancels the negative grid potential. These ions produce grid current, and a protective series resistance must be inserted into the circuit to limit the current to a safe value. Up to the moment of firing there are no positive ions, practically no current flows to the negative grid, and firing is controlled with almost no energy expenditure.

The thyratron can be stopped by first making the grid more negative than the critical value and then momentarily removing the plate voltage. Making the grid negative does not of itself stop the current flow, but it puts the grid in a position to control the flow once the positive ions disappear. Removing the plate voltage drops the plate current to zero, and after an instant called the deionizing time, the positive ions recombine with electrons to produce normal gas atoms. The negative grid then regains control, despite reapplication of the plate voltage. The deionizing time ranges from about 1,000 microseconds for most mercury-vapor thyratrons to less than 100 microseconds for some types.

Figure 5.12 shows curves for a typical gas-filled thyratron of the receiving-tube size suitable for control applications involving small currents. The characteristics are relatively independent of the temperature, and the tube may be used at low temperatures without danger of increased tube drop and cathode disintegration. Large thyratrons capable of handling several amperes or more usually contain mercury vapor and need to be protected by automatic controls to provide proper operating temperature. Their control characteristics also vary considerably with the condensed mercury temperature.

Shield-grid Thyratrons. To increase the control sensitivity, many thyratrons are of the shield-grid type, illustrated by Fig. 5.14.

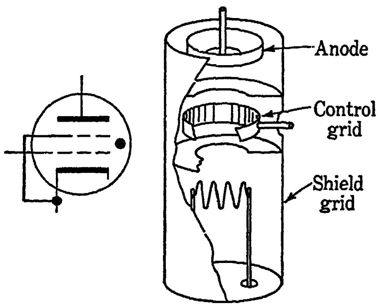


FIG. 5.14. Simplified sketch illustrating the construction of a shield-grid thyratron.

The shield grid, at approximately cathode potential, is designed to provide the majority of the electrostatic shielding between plate and cathode, but not enough to prevent plate-current flow with a moderately positive anode voltage. The control grid, in the form of a ring around the current-flow path, provides the remainder of the controlling field and produces control characteristics similar to that of the triode type. This construction has the advantage of reducing the control-grid area and placing it outside the path of current flow. This reduces both the grid current and the control-grid-anode capacitance. Lower grid current represents less control power, and such tubes are used in applications where the control signal source has a high impedance. In other respects the shield-grid thyratron compares with the triode type.

5.7 A-C Control of Thyratrons

The problem of controlling both the starting and stopping of a thyratron can be solved neatly by using an alternating voltage in the plate circuit. To investigate this possibility, let us analyze the performance of the circuit of Fig. 5.15, which represents the basic element of many thyratron control circuits. Voltage E_c represents a control voltage which might come from a photo-

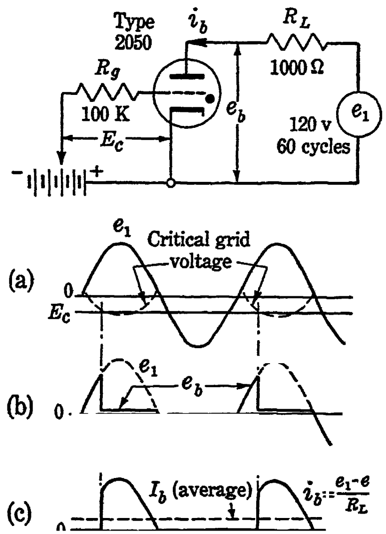


FIG. 5.15. Basic thyratron control circuit with an a-c plate supply.

electric tube circuit, R_g is a resistor to limit the grid current as specified by the manufacturer, and R_L is a load resistor to limit the plate current after the tube fires. In practice, R_L could be a relay to be closed or a lamp to be lighted. The a-c source might be the 120-volt line, as shown, or a transformer for higher voltages.

Figure 5.15*a* shows the wave form of the applied alternating voltage e_1 and the corresponding critical grid voltage curve to an enlarged scale. This dotted curve is plotted, point by point, by reading e_1 for various time moments and determining corresponding values of critical grid voltage from the control characteristic of Fig. 5.12. The curve obtained represents conditions up to the moment of firing, since before that time no plate current flows through R_L, and e_1 equals the plate voltage on the tube.

Let us suppose that the grid voltage E_c is negative and intersects the critical curve, as shown in Fig. 5.15*a*. Up to the moment of intersection the grid is more negative than critical value and no current flows, but at the moment of intersection ionization takes place and the grid loses control. As soon as the tube fires, the plate voltage e_b across it drops to a low constant value (Fig. 5.15*b*) and the plate current is determined by the values of e_1 and R_L for the remainder of the cycle. This produces a plate-current pulse looking like a portion of a half sine wave, as shown by Fig. 5.15*c*.

At the end of the half cycle, when e_1 drops below the ionizing potential of the tube, conduction stops and the thyratron deionizes during the negative loop of the supply voltage. This leaves it ready to repeat the performance cycle after cycle. The plate current pulses can be smoothed out if necessary.

With a sufficiently negative grid bias, no intersection occurs at any time and the tube remains completely nonconducting, as indicated by Fig. 5.16*a*. At some critical point (Fig. 5.16*b*) the tube

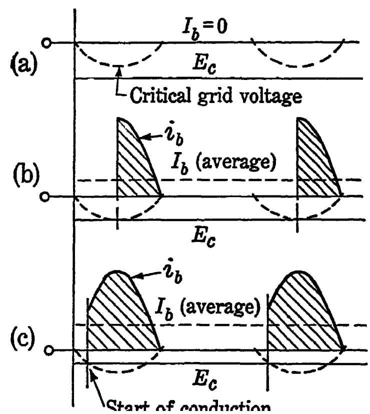


FIG. 5.16. Control of the average thyratron plate current by grid-bias adjustment.

just starts to conduct for approximately one-quarter of a cycle. For zero grid bias the tube conducts for full half cycles with twice the average value of plate current obtained under the critical condition of Fig. 5.16*b*. These diagrams show that an alternating supply makes it possible to start and stop the tube by controlling the grid bias only, with the limitation that the tube can start only

during positive half cycles with continued conduction to the end of a given half cycle. For many purposes this gives close enough time control, but a "finer grained" control can be obtained by increasing the supply frequency.

5.8 The Mercury-arc Rectifier

The term mercury-arc rectifier describes an arc-discharge diode with a mercury-pool cathode. In practice, two or more such diodes are enclosed in a single envelope with a common pool cathode. This construction fits the requirements of many of the full-wave rectifier circuits of Chap. 3 and of the polyphase circuits discussed in Chap. 8, which call for a number of diodes with the cathodes connected together. Large mercury-arc rectifiers usually consist of a water-jacketed steel tank with a recess in the bottom for the mercury pool. The graphite anodes are supported by insulating bushings from the tank top. Additional details include some type of ignitor, excitation anodes for maintaining the arc, and internal cooling coils for condensing the evaporated mercury. The whole assembly must be mounted on insulated supports because the tank and cathode are normally from several hundred to several thousand volts above ground potential.

Mercury-pool Cathode. The emitting surface in a mercury arc consists of a small bright cathode spot moving rapidly around on the mercury surface. This spot emits electrons which produce ionization as they speed to the anode; the ions, in turn, maintain the spot by cathode bombardment. For currents above about 20 amperes, the spot divides and subdivides until enough spots exist to provide the emission demanded. Mercury evaporated by the cathode spot condenses on the cool envelope walls and returns to the pool. Nearly a gram of mercury is evaporated for each 100 ampere-seconds of current flow. The vapor pressure due to this mercury must be limited by providing sufficient cool surface for condensation. Large rectifiers employ water cooling, both for this purpose and to remove the energy lost in the arc.

This type of cathode has the advantage of ruggedness and ability to provide high emission currents on demand. Practically the only limit to the current flow is the ability of the cooling system to dissipate the energy loss and to reduce the vapor pressure to maintain a reasonably high inverse voltage. This is in decided

contrast to the thermionic gas diode, with its need for careful temperature control and its delicate oxide-coated cathode unable to stand high overloads without disintegration.

Arc Ignition. The chief disadvantage of the mercury arc is failure of the arc to start automatically when the plate voltage reaches the ionizing potential. Instead, the arc must be initiated by mechanically breaking a current-carrying contact with the mercury pool, or by using a pulse of current through a stationary ignitor. The mechanical system is used in mercury-arc rectifiers, and the stationary ignitor is employed in the *ignitron* discussed later in this chapter.

Figure 5.17 shows a full-wave single-phase mercury-arc rectifier

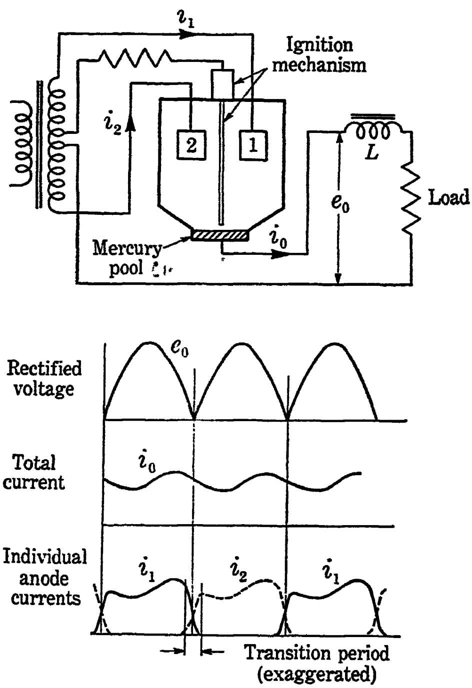


FIG. 5.17. Single-phase full-wave mercury-arc rectifier with inductance filter. The wave forms show that the cathode current is continuous to maintain the cathode spot.

to illustrate the principle of ignition and arc maintenance. When the supply power is applied, no arc starts until the ignition device

operates. In some cases this consists of a metal rod operated by an electromagnet and connected to an auxiliary source of voltage. The rod is first dipped into the mercury pool and then withdrawn. When the contact breaks, local heating initiates a cathode spot and the emitted electrons immediately drawn to a positive anode produce the ionization required to maintain the spot. In a half-wave rectifier, current would flow until the end of the positive half cycle and then the arc would go out. In a full-wave circuit with a series inductance, however, the arc can be maintained because the inductance demands a continuous current flow from the system, as shown by the wave forms of Fig. 5.17. The length of the current transition period during which the current flow changes from one anode to another has been exaggerated for clarity. This shows that for a brief time both anodes conduct, the cathode current is continuous, and the cathode spot is maintained.

The problem in polyphase rectifiers is simpler because at no time does the current ever reach zero, even without the assistance of a series inductor. It is possible, however, for the load current to drop to zero momentarily and extinguish the arc. This can be avoided by supplying a small fixed load, or by using auxiliary excitation electrodes connected to a low-voltage transformer to keep the power loss small.

The ignition system indicated by Fig. 5.17 can be made automatic by using the load current to withhold the igniting rod. A two-winding solenoid operated by both the load current and the ignition current suspends the rod until such time as the load current drops to zero. When the rod drops into the pool, the current flow in the ignition circuit again withdraws the rod and initiates the cathode spot. If the arc is taken up by the main anodes, the rod is then withheld until the next time ignition is necessary. If not, the operation repeats until the arc does strike.

Arc Back. One of the chief difficulties with mercury-arc rectifiers is the tendency to arc back. By arc back is meant the flow of current from one anode to another or a reverse flow from anode to cathode. This problem arises with any gas tube because the presence of the gas permits a reverse-current discharge to take place whenever the peak inverse voltage exceeds the sparking potential, but in a mercury-arc rectifier the situation is complicated by the continuous presence of positive ions in the tank. These positive ions may be attracted toward a negative anode to form

a cathode spot on it and permit reverse current flow. If, for instance, negative anode No. 2 of Fig. 5.17 should form a cathode spot, current flow between it and positive anode No. 1 would practically short-circuit the transformer.

It has been found that the installation of baffles surrounding and separating the anodes decreases the frequency of arc backs to a low value. Unfortunately, this also increases the arc drop to 20 volts or more instead of the 12 to 15 volts typical of an unbaffled arc. There is also the obvious disadvantage of having all the arcs in one basket; in case one portion of the rectifier needs repair, the whole circuit must be shut down.

5.9 The Ignitron

The ignitron is a mercury-arc tube in which a stationary ignition electrode starts the discharge at the beginning of each conduction cycle. This tube possesses all the efficiency and ruggedness of the mercury-arc rectifier plus many of the control features of the thyratron; for many purposes it has superseded the mercury-arc rectifier.

Ignition. Figure 5.18 shows a typical ignitron to consist of a steel water jacket enclosing an evacuated chamber containing a

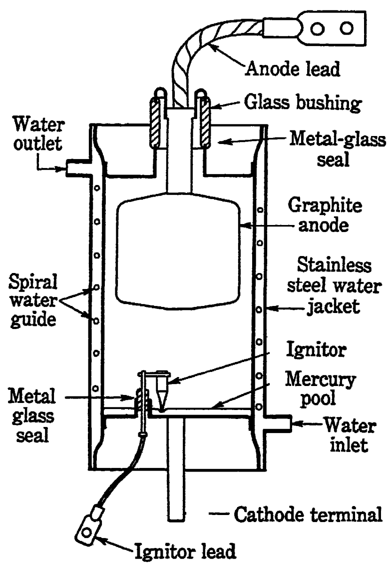


FIG. 5.18. Sectional view of an ignitron to show the component parts.

graphite anode, a mercury-pool cathode, and an ignitor. The pointed ignitor, about the size of a sharp pencil, consists of a highly refractory but conducting material such as silicon carbide. An insulated mount supports the ignitor with the point intersecting the mercury surface. A current pulse of about 20 amperes passed through the ignitor-mercury contact produces local heating and forms a cathode spot. The arc then immediately transfers to the positive anode, and ionization maintains the cathode spot for the remainder of the positive half cycle. At the beginning of each cycle a new ignitor-current pulse starts the arc. Pulse delay permits starting the ignitron at any point in the positive cycle to control the average value of the plate current. Although each tube requires auxiliary equipment to provide the ignitor pulse, the other advantages outweigh the additional circuit complications.

Advantages. The ignitron possesses three main advantages over the mercury-arc rectifier: (1) each tube contains only one anode, thus reducing arc-back difficulties to a minimum, (2) a separate tube is used for each diode in the circuit so that only one element has to be replaced in case of failure, and (3) the length of the current pulse can be controlled by the ignition circuit. With only one anode, the ionization dies down rapidly at the end of the conduction period and very little baffle is necessary to prevent arc back. This makes the tube drop as low as 12 volts—nearly as low as for the thermionic gas diode, yet without requiring the expenditure of external cathode heating power.

Ignitron Rectifier. Figure 5.19 shows the circuit diagram for a

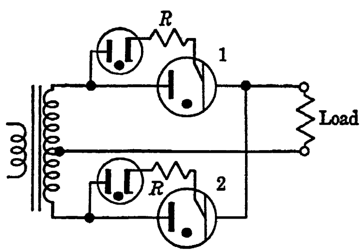


FIG. 5.19. Single-phase full-wave ignitron rectifier with mercury-vapor diodes for providing the ignitor current. Ignitrons are ordinarily used in polyphase rectifiers.

single-phase ignitron rectifier employing two mercury-vapor diodes to operate the ignitors. Let us suppose that tube 2 has been

conducting and that tube 1 is about to start. When the anode of tube 1 becomes positive, it cannot immediately conduct because there is no cathode spot. Instead, conduction takes place through the parallel ignition circuit until the current has reached a sufficiently high value to cause ignition. At this moment the voltage across the ignitron equals the mercury-vapor tube drop plus the voltage drop in R and the ignitor. As soon as the ignitron fires, the total voltage drop decreases to the arc potential of the ignitron, which is about the same as for the mercury-vapor diode, and the current through the ignition circuit drops to practically zero. Of course this system is effective only with large rectifiers handling big load currents compared with the required ignitor current.

Resistance welding is another important application of the control and peak current capabilities of the ignitron.

PROBLEMS

5.1 A type OC3 voltage-regulator tube is connected into the circuit of Fig. 5.6 with an input voltage of 250 volts and a load drawing 50 ma. Compute (*a*) the value of R required to permit the tube to operate at 20 ma, and (*b*) for this value of R the maximum and minimum allowable input voltage to keep the glow tube within its rated limits of 5 to 40 ma.

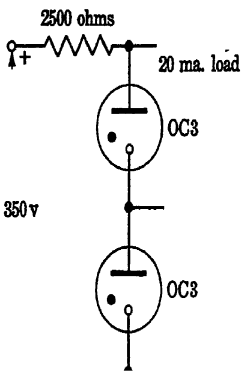


FIG. 5.20. Regulated voltage supply.

5.2 The circuit of Fig. 5.20 provides two regulated voltages from a single power supply. Determine the maximum current that can be drawn by a load connected between the mid tap and the negative terminal without causing either glow tube to operate outside its rated current range.

5.3 A type 2050 thyratron operates in the circuit of Fig. 5.15. (*a*) Draw accurately to scale the critical grid voltage line for a tube having the average characteristics of Fig. 5.12. (*b*) Compute the peak instantaneous anode current (tube drop is 8 volts). (*c*) Compute the average plate current for a bias adjusted to allow the tube to fire for one-quarter of a cycle.

5.4 The thyratron circuit of Prob. 5.3 operates with an alternating grid voltage of 5 volts effective lagging the plate supply by 60 deg. Graphically determine the angle at which conduction starts.

CHAPTER 6

PHOTOSENSITIVE DEVICES

PHOTOSENSITIVE devices can be divided into three classes: (1) photoemissive tubes which depend on the emission of electrons from a photocathode, (2) photovoltaic cells which produce a small generated voltage when exposed to light, and (3) photoconductive cells which change resistance when illuminated. Photoconductive cells have not been developed commercially to the same extent as the photoemissive and photovoltaic types.

6.1 Photometric Terminology

Before discussing the characteristics of the various types of photosensitive devices, it is well to review some of the photometric terms employed. These are given in English units because commercial phototubes are rated in this system.

International Candle. The unit of luminous intensity is the standard candle consisting of a certain type of lamp operated under standard conditions. One uniform standard candle emits 4π lumens.

Luminous Flux. Luminous flux refers to the rate of flow of radiant energy evaluated with respect to its visual sensation. Thus a microwatt of ultraviolet radiation represents no luminous flux, whereas the same energy in the visible spectrum does.

Lumen. The lumen is the unit of luminous flux. It equals the flux emitted in unit solid angle from a uniform source of 1 international candle. It has the dimensions of energy per unit time or power. For tubes operating in the infrared or ultraviolet region the lumen has no particular significance, and the ratings are often based on microwatts of incident energy. For a given frequency this is proportional to the number of photons per second.

Foot-candle. The standard of illumination is the foot-candle, which is the illumination on a surface placed 1 foot from a standard

candle. In general, the illumination is defined as the density of the luminous flux incident at that point. For a uniformly illuminated surface this means that

$$\text{Foot-candles} = \frac{\text{lumens}}{\text{square feet}} \tag{6.1}$$

This has the dimensions of power per unit area and is proportional to the number of photons per second per unit area.

6.2 Photoemissive Tubes

A photoemissive tube consists of an anode and a photosensitive cathode enclosed in a transparent envelope. Figure 6.1 shows a typical construction; the cathode is large to intercept the incident light, and the anode is small—often only a thin wire. The tube may be highly evacuated, or it may be gas-filled to increase the sensitivity. Photons striking the cathode liberate photoelectrons which are attracted by the positive anode. Depending upon the characteristics of the cathode, 1 lumen may produce an emission current of between 10 and 50 microamperes.

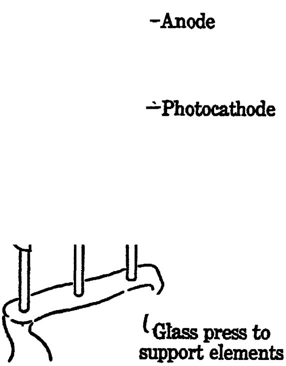


Fig. 6.1. The internal construction of a photoemissive tube.

Spectral Sensitivity. The spectral sensitivity of the tube can be controlled by the processing of the cathode surface. One common type of cathode consists of oxidized silver coated with a monatomic layer of cesium. This surface possesses a very low work function and provides relatively large emission currents. Figure 6.2 shows several standard types of spectral sensitivity characteristics useful for different applications. The S_1 curve provides high sensitivity to tungsten light, which is strong in red and infrared radiations. The S_3 surface is sensitive in the visible region with especially high sensitivity in the blue and near ultraviolet. The ultraviolet sensitivity in these tubes is limited by the transmission of the glass envelope. The S_6 curve is obtained by combining an ultraviolet sensitive cathode with a special envelope having an extremely thin window to admit the light. Such a tube is used for the laboratory measurement of ultraviolet radiation.

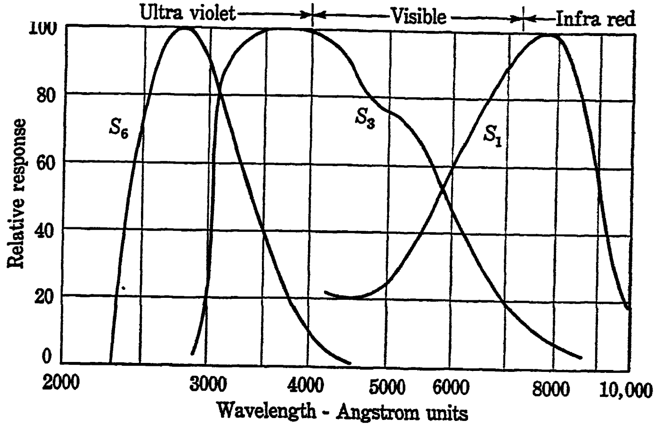


FIG. 6.2. Standard spectral sensitivity characteristics for several different photosensitive surfaces.

and 6.2

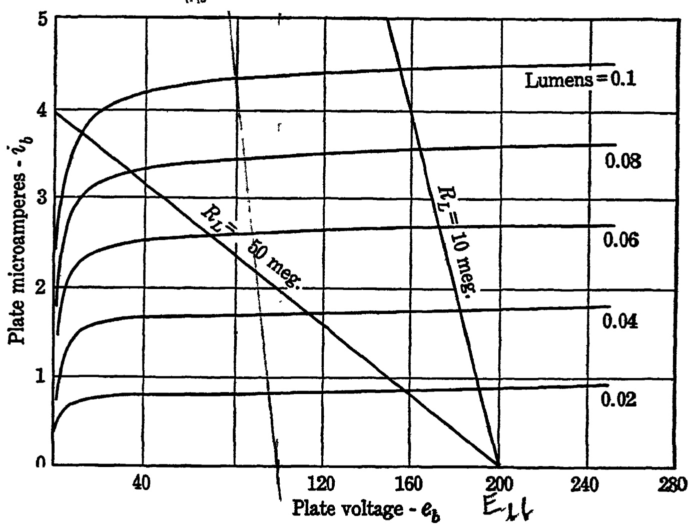


FIG. 6.3. Average plate characteristics for a type 929 vacuum phototube.

Vacuum Phototube. The electrical behavior of a vacuum phototube can be shown by a family of characteristic curves like those of Fig. 6.3. Each of the curves for a constant value of light input was obtained by varying the plate-supply voltage and taking simultaneous readings of voltage and current. The tiny currents

produce so little space charge that each curve rises rapidly to the saturation point and then levels off. In the saturated region, where all the photoelectrons reach the anode, the current is accurately proportional to the number of photons arriving per second and thus also proportional to the incident lumens. This linearity is maintained for all light intensities up to the point where the intensity of radiation is sufficient to damage or change the properties of the cathode. Most phototubes should not be exposed to illumination levels exceeding several hundred foot-candles even when disconnected from any circuit. Artificial lighting seldom reaches these values, but strong daylight or direct sunlight may injure the cathode.

Although the phototube currents are too small to operate a normal relay, it is possible to obtain relatively large voltage drops by passing the current through a high resistance. Figure 6.4

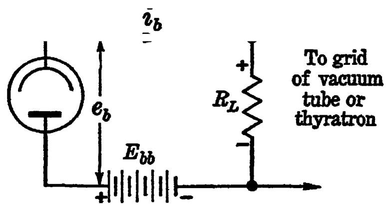


FIG. 6.4. A phototube circuit which produces a voltage drop in R_L proportional to the light intensity.

shows such a basic phototube circuit with a supply battery E_{bb} for maintaining the anode positive and resistor R_L across which the voltage drop is produced. This voltage drop, in turn, operates a vacuum-tube amplifier or thyratron control circuit.

The voltage drop across the phototube equals the supply voltage minus the drop in resistor R_L. Expressed mathematically,

$$e_b = E_{bb} - i_bR_L \tag{6.2}$$

This is the equation of a load line drawn in exactly the same fashion as for the vacuum triode. Figure 6.3 shows several such lines. Although they all represent rather high resistances, they are not too high for use in the grid circuit of a vacuum tube or a shield-grid thyratron. With a 10-megohm resistor, for example, an increase of light flux amounting to 0.1 lumen will produce a 45-volt change across R_L. To produce a 5-volt change sufficient to control the grid of a type 2050 thyratron (Fig. 5.12) requires an illumina-

tion change of only 0.011 lumen. This corresponds to a change of 1.6 foot-candles on a cathode having an effective area of 1 square inch.

At high light intensities, where the load-line intersections extend into the curved region of Fig. 6.3, the voltage changes become nonlinear. If linearity is important, the load resistance must be low enough to avoid this possibility.

The operation of a vacuum phototube is practically instantaneous because of the extremely short electron transit time and low anode-cathode capacitance. This is not true of the other types of photoelectric devices, and their field of use is correspondingly limited.

Gas-filled Phototube. The phototube sensitivity can be increased by filling the envelope with a low-pressure inert gas to take ad-

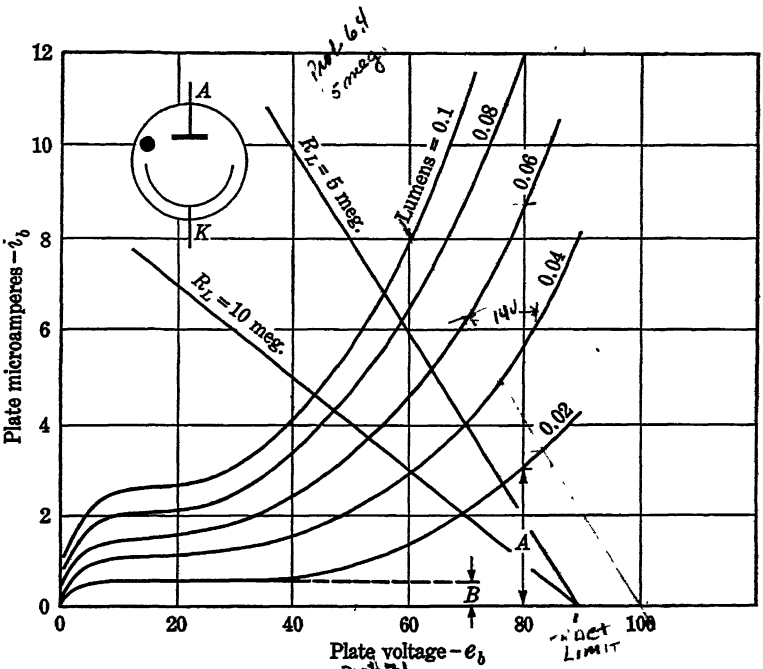


FIG. 6.5. Average plate characteristics for a type 930 gas-filled phototube.

vantage of the current amplification produced by the ionizing electron avalanches of the Townsend discharge. Figure 6.5 shows characteristic curves for this type of tube. Each curve shows the

typical level region followed by a rising portion indicative of ionization by collisions between electrons and gas atoms. At high applied voltages the current rises rapidly, and the discharge may pass over into the glow form. For this reason the applied voltage must be carefully limited to the rated value, usually about 100 volts.

Inspection of the curves shows the intercepts on the load line to be nonuniform, but the current changes are so much larger than with a vacuum tube that a given output voltage change can be obtained with either a lower load resistor or with smaller illumination changes. For off-on relay service, linearity is of little importance anyway.

At low light levels, where operation takes place near the bottom end of the load line, the linearity is not bad because the phototube drop is little changed by load-line excursions and the effective amplification of the photoemission by ionization is practically independent of the current. This increase is called *gas amplification*, and the ratio between the initial photoemission and the final anode current is called the *gas-amplification factor*. Figure 6.5 indicates how this factor can be determined from the curves. Height B is taken from the level portion of the lowest characteristic curve and measures the original photoemission at a voltage too low to produce any ionization. Height A, measured on the actual curve for the voltage at which the gas amplification is desired, represents the photoemitted current plus the additional ionization current. The gas amplification is defined as the ratio A/B—about 5 for the example shown. At low light levels a gas phototube acts about like the corresponding vacuum tube, with all the current values multiplied by the gas-amplification factor.

The gas-filled phototube is slightly slower in action than the vacuum type because of the relatively low velocities of the positive ions. This makes its *dynamic* response to rapidly varying light intensities less than its *static* response to a steady light source. Up to a frequency of 10,000 cycles per second the difference is not important, and gas phototubes are often used in the reproduction of sound on film.

6.3 Photovoltaic Cells

A photovoltaic device is one that directly converts radiant energy into electrical form. It not only provides a current proportional

to the illumination, but it also produces a small electromotive force capable of forcing this current through a low-resistance circuit. In this respect it is essentially different from the photoemissive cell, which requires a source of voltage to attract the emitted electrons. Photovoltaic cells also have the advantage of providing current of the order of a hundred microamperes; this is sufficient to operate directly a sensitive meter or relay without intervening amplification.

Commercially available photovoltaic cells are either of the copper-oxide or selenium barrier-layer type, closely related in construction and theory to the disk rectifier discussed in Chap. 2. To illustrate the operation of such a cell, let us consider a simple copper-oxide photocell as shown by Fig. 6.6. In many respects

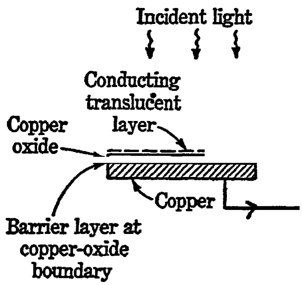


FIG. 6.6. Section through a copper-oxide photovoltaic cell. Compare this diagram with that of Fig. 2.10 showing a copper-oxide contact rectifier.

this cell is similar to the copper-oxide disk rectifier of Fig. 2.10, except that the front contact is made by a translucent conducting coating instead of a lead washer. This coating may be a thin sputtered metal layer, a transparent conducting lacquer, or a fine metallic grid in contact with the surface. The resulting sandwich is a barrier-layer rectifier in which the forward direction of conventional current flow is from the oxide to the copper, as discussed in Chap. 2. In the illustration the thickness of the various layers is greatly exaggerated.

When light shines on the cell, it penetrates the translucent front surface and the photons release electrons in the copper-oxide layer. As in the photoemissive tube, the number of photoelectrons produced per second is directly proportional to the light intensity. Furthermore, these photoelectrons are given kinetic energy by the photons; if the cell is connected to an external cir-

cuit, the electrons may be able to continue their motion and produce a current. Since the directions of the photoelectron velocities are random, current might be expected to flow equally in either direction with zero net effect. However, the cell is a rectifier, and its resistance to flow one way is greater than in the reverse direction; so it is not surprising to find that the majority of the electron flow is in one direction, and a meter connected to the cell registers a current. The disturbing fact is that the actual flow is opposite to the normal current direction when used as a rectifier. This is difficult to explain; in fact the whole theory of the barrier-layer interface is not well understood.

The photovoltaic cell of Fig. 6.6 is called a *back-effect* cell because the barrier is at the back of the oxide coating. By suitable processing a barrier layer can be built up between the translucent front layer and the oxide. In this case both the rectifying properties

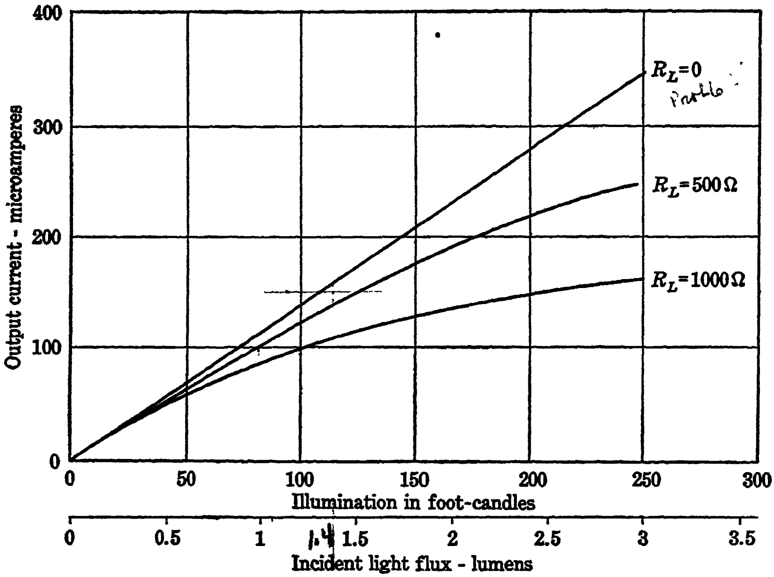


FIG. 6.7. Load-current curves for a Weston Photronic type of photovoltaic cell.

and the direction of the photoelectric current are reversed. This is called a *front-effect* cell; the ability to produce this type of cell testifies to the original random direction of motion of the photoelectrons.

Cells can also be constructed using iron and selenium, and iron

and iron selenide. Their construction and general properties are similar to those of the copper-oxide cell.

Characteristic Curves. Photovoltaic cells are surprisingly efficient, and they provide currents many times larger than do photoemissive cells. Figure 6.7 shows the characteristic curves for a Weston Photronic cell of the iron-selenide type. These curves show the cell current as a function of the illumination for various values of external load resistance. For very low load resistance (usually the indicating meter itself) the current is accurately proportional to the illumination, as might be anticipated from the fact that the number of photoelectrons produced is proportional to the illumination. With a larger load resistance the linearity is not good and the curve is flatter at the higher light intensities, although at low intensities the current is independent of the external resistance. This suggests that the cell produces a current proportional to the light intensity but that the fraction of the electrons that can continue around the circuit depends upon the countervoltage developed across the load resistance. This is reasonable because the photoelectrons probably have a range of velocities—the voltage drop in the load resistance stops the slower ones.

Although the Photronic cell produces relatively large currents, the voltage output is very small. Figure 6.8 shows the open-

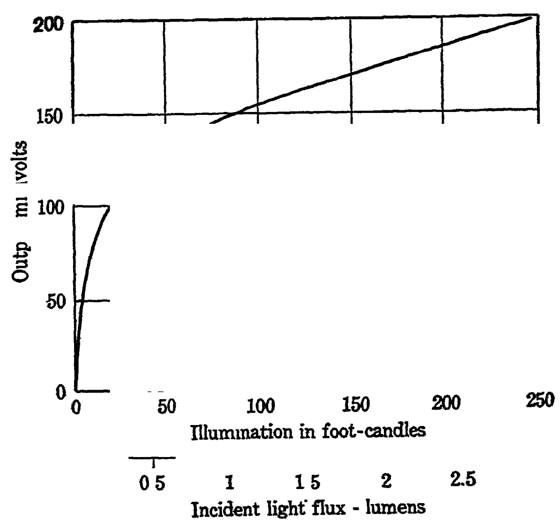


FIG. 6.8. A curve showing the open-circuit voltage generated by a Weston Photronic cell.

circuit generated voltage as a function of the illumination. This voltage was measured without drawing any current from the cell and represents the condition that might obtain if the cell were used to operate the grid of a vacuum tube. The curve shows that the voltage-current relationship is not linear and that the voltage changes are tiny compared with those obtained with the photoemissive tube circuit of Fig. 6.4. Therefore a photovoltaic cell is not particularly useful for operating a vacuum-tube amplifier or a thyratron control circuit.

The spectral sensitivity of the photovoltaic cell is similar to that of photoemissive tubes, and it is well suited to the measurement of visible light. Figure 6.9 shows the relative sensitivity

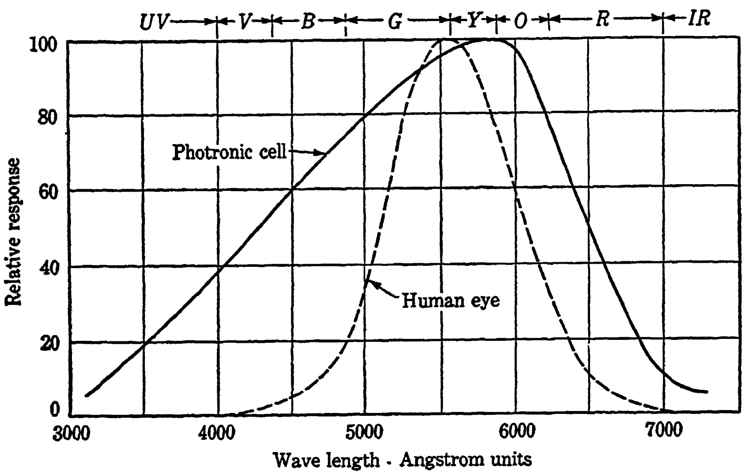


FIG. 6.9. A comparison of the relative spectral sensitivities of the human eye and the Photronic cell.

of the Photronic cell to light of different wavelengths as compared with the visibility curve of the human eye. With a greenish filter over the cell, the sensitivity curve can be altered to correspond exactly with that of the eye. This is necessary when the cell must accurately measure the visual effect of a light source.

Field of Use. Photovoltaic cells are perfectly suited for portable equipment to measure illumination. No batteries or amplifier are required, and the generated current is sufficient to operate a reasonably rugged meter. The common photographic exposure meter is an example of this application. These usually sacrifice

accuracy in the interest of compactness and portability. Better cells and meters are available for the accurate measurement of illumination level.

A photovoltaic cell can also operate an especially sensitive relay designed for this purpose. A relay of this type is rather delicate, and the contacts cannot handle much current; so it is often better to use a photoemissive tube controlling a thyratron.

A photovoltaic cell cannot be used for applications involving the observation of rapidly varying light intensities, such as the reproduction of sound on film. The sandwich construction with the two conducting electrodes separated by a thin film of poor conductivity produces a high internal capacitance that electrically shunts the output terminals. This causes the output to drop seriously for alternating light frequencies above about 100 cycles per second. For relay operation and illumination level measurements, this deficiency is unimportant.

6.4 Photoconductive Cells

A photoconductive device is one which changes its resistance in response to illumination. Selenium exhibits this property, and cells have been constructed that show considerable change in resistance when exposed to light. Such a cell must be used with an external source of voltage to produce either a voltage drop for amplification, or a current change to operate a meter or relay. They have not yet been developed commercially to any extent, but there is a future possibility that photoconductive cells may be capable of competing with the other classes of photosensitive devices.

PROBLEMS

6.1 Plot dynamic characteristic curves of plate current against lumens for a type 929 photocell for both load lines shown on Fig. 6.3.

6.2 The projected cathode area of a type 929 photocell is a rectangle $\frac{5}{8}$ by $\frac{13}{16}$ in. The cell operates in the circuit of Fig. 6.4 with $E_{bb} = 100$ volts and $R_L = 5$ megohms for the purpose of controlling a thyratron grid. If a 3-volt change is enough to control the thyratron, what is the minimum illumination change (in foot-candles) that can be observed by the circuit?

6.3 Repeat Prob. 6.1 for a type 930 gas-filled phototube.

6.4 Repeat Prob. 6.2 for a type 930 gas-filled phototube. This tube has the same cathode construction as the type 929 tube.

6.5 A sensitive relay designed for operation with photovoltaic cells has a resistance of 200 ohms and closes with a current of 150 μa. Determine the required light flux to close the relay with the cell of Fig. 6.7.

6.6 Compute and plot a curve showing the spectral transmission characteristic of a filter designed to make the response of a photronic cell fit that of a human eye. Assume that the filter transmits 90 percent of the incident light at a wavelength of 5,500 A.

6.7 A photovoltaic cell having the characteristics of Fig. 6.7 operates a 500-ohm meter with a full-scale reading of 100 μa. At what distance should the cell be placed from a 100-cp lamp in order to produce full-scale deflection of the meter?

CHAPTER 7

ELECTRONIC CONTROL CIRCUITS

THE PURPOSE of this chapter is to discuss a few elementary types of electronic control circuits, to suggest the possibilities of such circuits, and to provide exercise in understanding how they operate. The control possibilities of electronic and electromechanical equipments seem to be unlimited; they measure time, count operations, observe color, brilliance, and position, perform mathematical operations, control speed, remember selected facts, and open doors at our approach. However, an electronic circuit may not necessarily do the best or the simplest job. It is not always true that "You can do it better with tubes." Mechanical elements may be simpler and more reliable, and many control circuits involve both electrical and mechanical parts.

7.1 The Relay

One of the simplest electromechanical devices is the relay shown by the conventional diagrams of Fig. 7.1. The relay consists of an

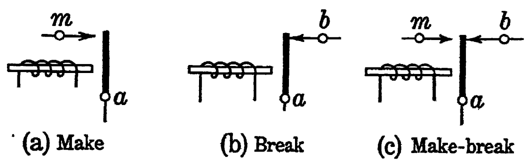


FIG. 7.1. Conventional representation for several basic relay types.

iron core wound with a number of turns of wire. Current flowing through this coil attracts the iron armature a and moves a number of electrical contactors. A restoring spring returns the armature to its normal position when the current is turned off. The diagram shows all the relays in the *normal*, or nonenergized, position. Figure 7.1a shows a type of relay that closes or "makes" the contact when energized. Diagram b illustrates the type that

breaks the circuit when energized, and c shows one that amounts to a single-pole double-throw switch. The majority of simple relays are like c and can be used for make, break, or switching service. More complex relays can be obtained with almost any multiple combination of contacts intended for some specific purpose. Remote-control motor starters, dial-telephone systems, and circuit breakers are all examples of relay systems designed for specific purposes.

Relays are made in a wide variety of sizes and contact current ratings. Some of the most sensitive will close with an energizing current of a hundred microamperes or less. These are built like sensitive D'Arsonval meters, but with the pointer replaced by an arm carrying contacts capable of handling small currents. Larger, more rugged relays designed for currents of a few milliamperes have contacts that can safely carry an ampere or more. These relays usually have a fairly high resistance and are designed for use in the plate circuits of small receiving-type vacuum tubes and small thyratrons. Still larger relays, that close with control currents of 0.1 ampere or more, carry sufficiently heavy contacts to handle considerable power—enough to start a motor, perhaps.

The sensitivity of a relay is determined by the number of ampere turns carried by the coil rather than by the current alone. Consequently, a high-resistance coil of many turns of fine wire carrying 10 milliamperes is just as effective as a low-resistance coil with one-tenth the number of turns and carrying 100 milliamperes.

7.2 Thyratron Relay Circuit

One simple and useful circuit is the thyratron relay of Fig. 7.2. The purpose of this circuit is to control the flow of considerable power by means of a delicate contact that is itself incapable of handling appreciable current. The contact might be a fine wire feeler observing the deflection of a sample in a testing machine, or it could be a tiny wire insert in the column of a mercury thermometer to close the circuit at a definite temperature. Such tiny contacts usually make and break very slowly, and the contact resistance when closed is often variable and not particularly low.

Resistors R_1 and R_2 of the circuit are chosen to provide the proper heater current and to establish the cathode at some potential intermediate between the two terminals of the a-c supply voltage. The voltages shown on the diagram are representative. The same result can be obtained by providing proper transformer

taps and using a separate winding for the heater, but the circuit shown has the advantage of simplicity, although it wastes some power in R_1 and R_2.

During the positive half cycle of e_1 the grid-circuit voltage e_2 stays below the critical firing potential so that no conduction takes

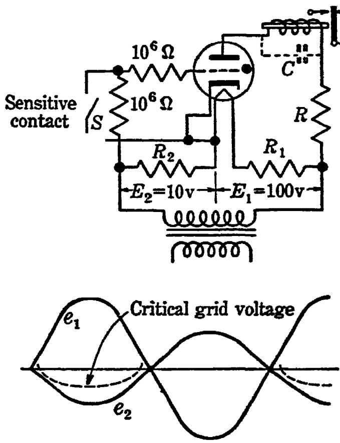


FIG. 7.2. Thyratron relay circuit. Closing the sensitive contact controls the main relay.

place. The grid current is also zero and the presence of the two 1-megohm resistors does not affect the grid voltage. During the next half cycle the plate is negative so that no plate current can flow. Grid current will flow because e_2 is positive, but it is limited to a very small value by the grid circuit resistance.

Closing the contact connects the grid directly to the cathode through a 1-megohm resistor. The grid is then no longer forced negative when e_1 is positive, and the tube conducts for full half cycles. This energizes the relay to perform whatever operation is desired. For a temperature-control circuit, closing of a mercury-column contact might operate the relay to open the electrical heating circuit producing the controlled temperature.

With S closed a small current flows through it equal in value to voltage E_2 divided by the 1-megohm resistor—in this case 10 microamperes. This can be reduced by increasing the circuit resistance to the limit allowed for the particular thyratron and by reducing voltage E_2 to the minimum necessary for control.

Resistor R plus the relay resistance must be sufficient to limit the thyratron plate current to the rated value, yet sufficient to

operate the relay. Since the plate current is in the form of pulses, the relay may tend to chatter if it is improperly designed. Condenser C across the relay can be used to smooth the relay current.

7.3 Photoelectric Control Circuit

The photoelectric control circuit of Fig. 7.3 is an adaptation of the circuit of Fig. 7.2, with the phototube taking the place of the

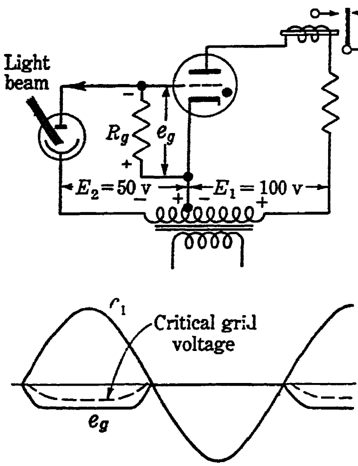


FIG. 7.3. Photoelectric control circuit. Darkness energizes the relay, and light releases it.

contactor S. For variety, a tapped transformer is shown; there must also be a low-voltage winding for heating the cathode. At a moment when the plate voltage e_1 is positive with respect to the cathode, the voltage e_2 applied to the photocell circuit is negative. As long as light reaches the cell, it can conduct through R_g to bias the grid below the critical voltage and prevent firing. During the reverse half cycle the phototube circuit is inoperative, but the negative plate voltage applied to the thyratron prevents conduction.

Interruption of the light beam drops the phototube current to a low value, and the lack of negative grid bias permits thyratron plate current to flow each positive half cycle. This closes the relay. Of course there is a critical value of light flux at which the phototube current produces a voltage wave e_g that just intersects the critical grid voltage line at its negative peak. Under this condition the tube will fire for one-quarter cycle and provide just

one-half the average current produced by full-cycle operation. By choosing the relay to close with this smaller current, the action of the circuit can be kept positive under all conditions. Light intensities higher than critical value will produce no action, and those below this value will definitely operate the relay.

This type of circuit is used in the familiar photoelectrically operated doors that open when a beam of light is broken by the person about to enter, and it can be used to count or eject articles moving on a conveyer belt.

For some applications it may be more convenient to reverse the process and have the relay energized by light and released during dark periods. This can be done by reversing the phototube and the polarity of the phototube voltage, as suggested by Fig. 7.4.

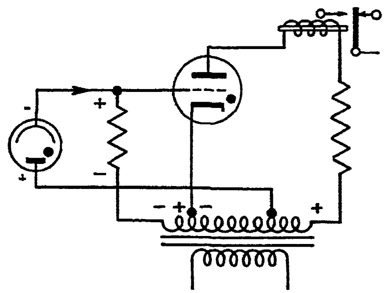


FIG. 7.4. This circuit performs in opposite fashion to that of Fig. 7.3. Here application of light energizes the relay.

7.4 Thyratron Time Delay

Figure 7.5 shows a simple and convenient circuit for producing an adjustable time delay. Closing switch S initiates the action,

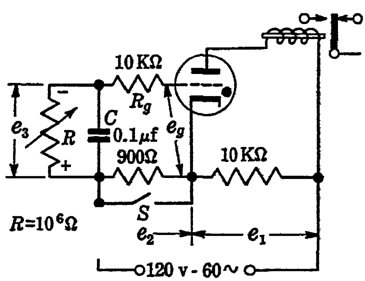


FIG. 7.5. Thyratron time-delay circuit. The relay closes a definite time after switch S is closed.

and after a predetermined time delay, the relay closes. The delay period is controlled by the R and C of the circuit; R is often an adjustable rheostat with a dial calibrated in delay time. One application of this circuit is on spot-welding machines, where a number of separate timers control the sequence of pressure, welding current, forging pressure, and release operations. All the delay timers are set into action simultaneously, with different delay settings to control the time sequence of the process.

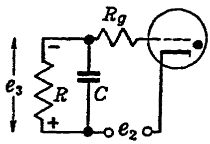


FIG. 7.6. With switch S open, the grid circuit of Fig. 7.5 becomes a half-wave rectifier with smoothing capacitor.

The operation of this circuit depends on the exponential discharge of an R–C combination to produce the delay. With switch S open, the grid circuit is similar to a half-wave rectifier with smoothing condenser, as suggested by Fig. 7.6. This produces a smooth rectified voltage e_3 approximately equal to the peak value of e_2, as shown in Fig. 7.7. The total grid-cathode voltage equals e_2 plus e_3, which is much more negative than the critical grid potential during the positive half-cycle of the plate-voltage wave.

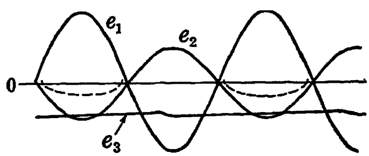


FIG. 7.7. Wave forms of Fig. 7.5 with switch S open. The grid circuit conducts to recharge C during the time that the plate voltage is negative.

Therefore the thyratron does not conduct, and the relay remains open. R_g serves to limit the size of the grid-current pulses.

Closing S removes voltage e_2 from the grid circuit, condenser C is no longer recharged each cycle, and e_3 decreases exponentially with time until the negative bias is no longer sufficient to keep the thyratron from firing. Figure 7.8 shows this gradual decay until the line representing e_3 finally intersects the critical grid voltage line at some particular cycle and the tube fires for a portion of a half cycle. During this first firing period the positive ion flow to the grid discharges C through R_g. This rapidly reduces the negative bias, and the thyratron fires for practically complete half cycles thereafter.

The delay time is proportional to the R–C product or time constant of the grid circuit. For Fig. 7.5 the R–C product is 0.1 second, and the delay obtained for the voltages given is about 0.2 second. By making R or C variable, the delay can be adjusted over a wide

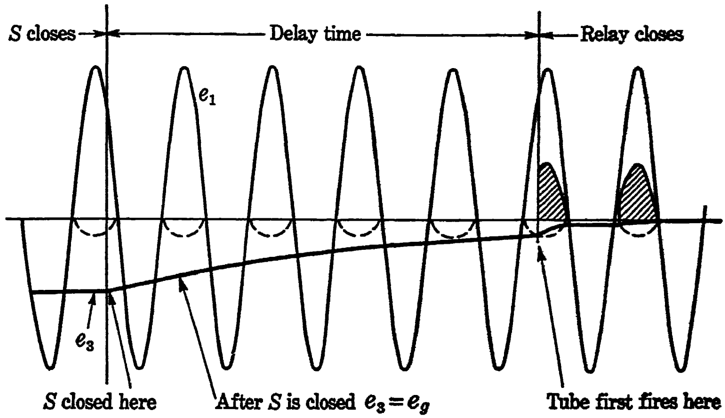


FIG. 7.8. Wave forms showing the operation of the thyratron delay timer. The heavy line shows the gradual voltage decay across C until it can no longer prevent thyratron conduction.

range up to a number of minutes if desired. For short delays the accuracy suffers from the fact that the control is not sufficiently "fine-grained" because the permissible delay times differ by discrete jumps from one cycle to the next and the period cannot be set closer than the nearest one-sixtieth of a second.

7.5 Thyratron Phase Control

Phase-shift control of thyratron conduction is useful in any number of applications where it is desired to produce a controllable rectified current. Figure 7.9 shows the phase-shift principle by which the thyratron firing point is controlled by adjusting the *phase* of an alternating grid voltage instead of varying its *amplitude*. Phase adjustment of the grid voltage permits moving the point of intersection with the critical grid voltage anywhere within the conducting half cycle and gives complete control over the average plate current I_b. This arrangement also makes the intersection of the grid-voltage curve with the critical voltage line less acute than is obtained with the amplitude control of Fig. 5.16 and therefore less affected by variations in the firing characteristic.

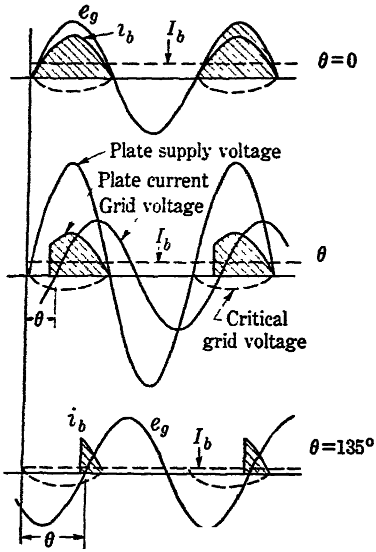


FIG. 7.9. Wave forms showing the effect of various amounts of phase shift on the average plate current of a thyratron.

Since the firing characteristics of thyratrons are none too reproducible, this is a considerable advantage.

After graphically determining the firing point for various values of phase shift, the average value of plate current as a function of angle θ can be computed. This gives a curve of the form of Fig. 7.10.

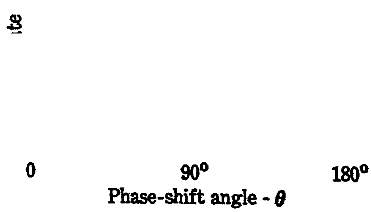


FIG. 7.10. This curve shows the effect of grid-voltage phase shift on the average thyratron plate current.

Phase-shifting Network. A simple phase-shifting network capable of providing a constant output voltage of variable phase position is shown by Fig. 7.11. A center-tapped transformer provides

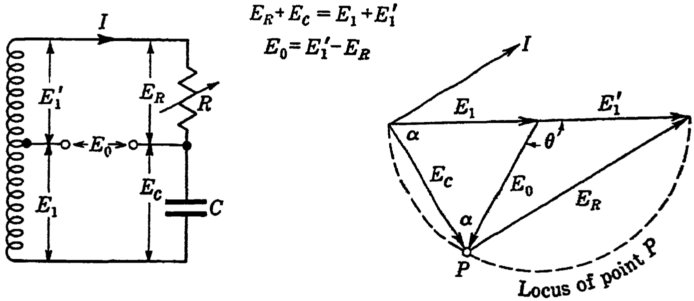


FIG. 7.11. Single-phase phase-shifting network. The vector diagram shows the effect of R and C on the output voltage E_0.

two equal voltages E_1 and E_1'. Connected in series across the transformer are a resistor R and capacitor C, both of which may be adjustable for controlling the phase of the output voltage E_0. It is assumed that no appreciable current will be drawn by the circuit utilizing the output voltage E_0.

The vector diagram is drawn with the total transformer voltage E_1 plus E_1' as the horizontal reference vector. Because of C, the current I must lead this voltage by an amount depending on the relative values of R and capacitive reactance X_c. A typical case is shown. Voltage E_r must be in phase with this current, as shown, and voltage E_c must lag by just 90 degrees. Also the sum of E_r and E_c must equal the total transformer voltage. Therefore vectors E_r, E_c, and the total transformer voltage form a right triangle, as indicated by the diagram. For other values of R and X_c the current I would lead the reference by a different amount, and E_r and E_c would form a different right triangle having a hypotenuse equal to E_1 plus E_1'. Since all triangles inscribed in a semicircle are right triangles, it follows that the locus of point P is a semicircle with the center located at the tip of vector E_1.

Referring to the circuit diagram, we see that the output voltage is

$$E_0 = E_1' - E_r$$

This corresponds to reversing the direction of the E_r vector and adding it to E_1', giving a vector E_0 as shown. Since E_0 extends to

point P, the dotted semicircle is also the locus of E_0. Consequently, varying the values of R and X_c changes the phase angle θ without affecting the magnitude of E_0.

Angle θ can easily be computed in terms of R and X_c. From the vector diagram,

$$\tan \alpha = \frac{E_r}{E_c} = \frac{IR}{IX_c} = \frac{R}{X_c}$$

$$\alpha = \tan^{-1} \frac{R}{X_c}$$

Phase angle θ is equal to just twice angle α. Therefore,

$$\theta = 2 \tan^{-1} \frac{R}{X_c} \tag{7.1}$$

Figure 7.12 shows the phase-shift network applied to the basic thyratron circuit for plate-current control. There are, of course,

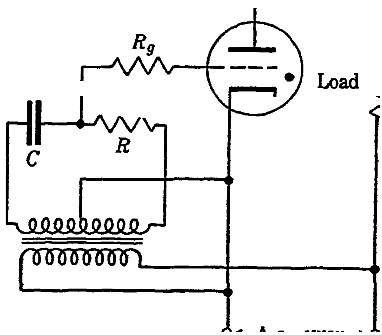


FIG. 7.12. An application of the phase-shifting network of Fig. 7.11 to a thyratron control circuit.

many other ways of producing a phase shift. With three-phase power the problem is simpler.

7.6 Thyratron Welder

Another common application of thyratrons and ignitrons is in resistance-welding apparatus. Thyratrons or thyratron-controlled ignitrons permit better control of timing and welding current than is possible with mechanical contactors, and they are less noisy. Their use is almost required for the welding of aluminum alloys because careful control of timing and welding temperature is necessary to produce uniformly strong welds.

The resistance-welding process consists of clamping together the two pieces of material to be welded and passing a sufficiently

heavy current through the contact to heat it to welding temperature. The combination of pressure and softening of the metal accomplishes the union. Speed is desirable to save time and to reduce the formation of interfering oxides. Sometimes the pieces to be joined are passed through the welder much like cloth through a sewing machine. After each weld the pressure is released, the material is stepped forward, pressure is reapplied, and a weld is made. Another system uses revolving pressure wheels and closely spaced welding pulses to produce a continuous seam.

Figure 7.13 shows the basic elements of a thyratron-controlled

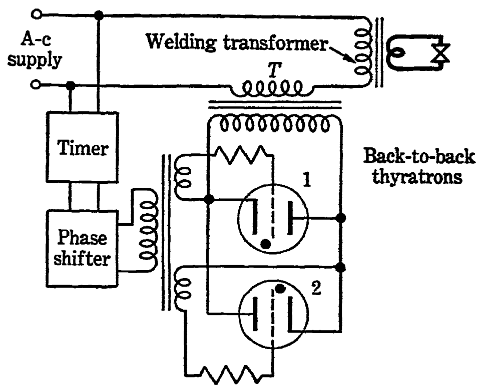


Fig. 7.13. Thyratron resistance welder with timing and phase-shift controls to control the duration and magnitude of the welding current.

resistance-welding system. The welding transformer is a high-ratio step-down transformer to provide large welding currents at low voltage. Transformer T is theoretically not necessary, but it permits the welding circuit to operate at higher current values than the thyratrons can carry directly. If, for instance, the lower winding of T has twice as many turns as the primary, the current in the thyratron circuit will be only half the line current. However, the alternating voltage impressed on the thyratrons will be twice the line voltage.

The two thyratrons are connected across the secondary of T in a "back-to-back" fashion. The alternating grid voltages are also connected to the two tubes in the opposite sense so that the two circuits are completely reversed in polarity in every respect. Positive-control thyratrons are employed for this particular circuit; the critical grid-voltage curve is in the positive region, and no conduction takes place without grid excitation.

The impedance presented to the welding circuit by the primary of transformer T depends on the load connected to its secondary. With the secondary open-circuited the only primary current flowing is the exciting current required for the transformer core; this is so small as to be negligible. If a short circuit is placed across the secondary, however, the impedance presented by the primary drops to a very low value and the welding transformer is effectively connected directly to the a-c line.

As long as the thyratrons are without excitation, they do not conduct and they appear as an open circuit to T. This practically disconnects the welding transformer from the line. When the timer applies grid excitation, each thyratron can fire for periods up to a full half cycle, depending on the phase-shift setting. With zero phase shift each tube fires for a full half cycle, and since they are connected with opposite polarity, they fire *alternate* half cycles. Tube 1 fires when the right-hand terminal of T is positive, and tube 2 conducts on the opposite polarity. The net effect is the same as though T were continuously shorted, and an approximately sinusoidal current flows through both its primary and

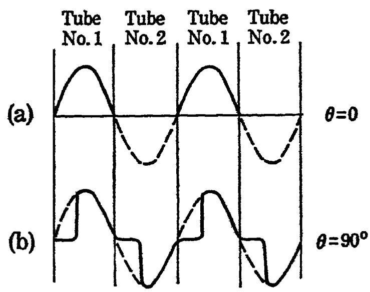


FIG. 7.14. Wave forms showing the effect of phase shift on the welding current produced by the circuit of Fig. 7.13.

secondary. Figure 7.14*a* shows this current and the contribution of each tube to it.

By shifting the phase of the grid voltage, the firing can be delayed to make each thyratron pass current less than half a cycle. The effective welding current is then reduced, as suggested by Fig. 7.14*b*. Phase shift thus gives a smooth control of the welding current from maximum to zero, and the timer permits selection of the number of cycles that current is allowed to flow in the weld.

Heavy-duty welders often employ ignitrons instead of thyratrons. The ignitrons are connected in a back-to-back circuit

without transformer T, and their firing is controlled by a pair of thyratrons or some other type of pulsing circuit.

7.7 Saw-tooth Generator

An interesting application of the thyratron is in circuits for the generation of saw-tooth wave forms. This wave form is used in cathode-ray oscillographs to move the spot repeatedly across the screen at a known rate. Figure 7.16 shows the circuit to be analyzed. The thyratron is a small one designed for this service. Both R and C are adjustable to control the saw-tooth frequency, and E_{bb} is provided by the supply rectifier.

FIG. 7.15. A saw-tooth wave form.

For any given grid-bias voltage, there is a critical plate potential

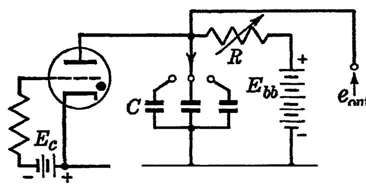


FIG. 7.16. Thyratron saw-tooth generator. This circuit is used in the majority of simple cathode-ray oscilloscopes to provide the horizontal deflecting voltage.

above which the thyratron will fire. This is shown by a horizontal line through point a on the diagram of Fig. 7.17. When the

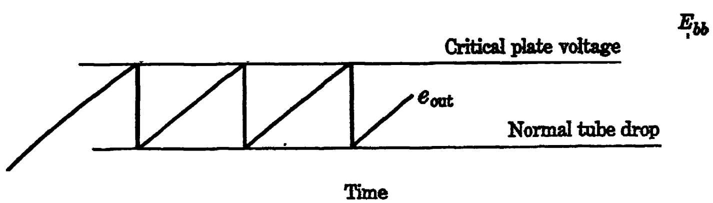


FIG. 7.17. This illustrates the generation of an approximately linear saw-tooth wave by the circuit of Fig. 7.16.

voltage E_{bb} is first applied, the thyratron is nonconducting and has no effect on the circuit. The current flow through R is spent in

charging C, and the voltage across it rises toward E_{bb} along the exponential curve. When the voltage reaches point a, the thyratron fires and practically short-circuits the condenser. This rapidly lowers the voltage to the normal tube drop indicated by point b on the lower line.

As soon as the condenser has discharged through the tube, it no longer contributes any current and the only path is through resistor R. With R sufficiently large, this current can be kept so low that the grid regains control and the tube again becomes nonconducting. The capacitor then starts to recharge, and the process repeats itself. The wave form is not perfect because the rise is slightly curved, but the wave can be made sufficiently linear for most purposes by using a high value of E_{bb} compared with the critical plate voltage.

PROBLEMS

7.1 The photoelectric control circuit of Fig. 7.3 operates with a type 930 gas-filled phototube and a type 2050 thyratron. Resistance R_g equals 1 megohm, and the incident light is 0.04 lumen. The effective alternating voltages are $E_1 = 100$ volts, and $E_2 = 50$ volts. (*a*) Plot the wave form of the voltage developed across R_g. (*b*) Determine if the incident light is enough to prevent conduction of *any* type 2050 thyratron whose characteristics fall within the normal range.

7.2 The circuit of Fig. 7.4 employs a type 2050 thyratron with 100 volts effective in the plate circuit. How large an alternating voltage must be introduced into the grid circuit to prevent thyratron conduction while the phototube is dark?

7.3 The phase-shift network of Fig. 7.11 employs a 1-μf capacitor and a variable resistor. Compute the range that must be covered by the resistor to permit phase-angle adjustment between zero and 120 deg with a 60-cycle input frequency.

7.4 (*a*) Derive the expression for the saw-tooth frequency of the circuit of Fig. 7.16 in terms of the tube drop E_d, the critical plate voltage E_k, E_{bb}, and the circuit constants. The charging voltage of a capacitor follows a curve similar to the discharge; $e = E(1 - e^{-t/RC})$. (*b*) Given $E_{bb} = 300$, $E_d = 10$, $E_k = 50$, and $R = 1$ megohm, compute the value of C required to produce a frequency of 1,000 cps.

CHAPTER 8

POLYPHASE RECTIFIERS

THE SINGLE-phase rectifiers discussed in Chap. 3 are commonly used for providing small amounts of d-c power, but large industrial rectifier units ordinarily operate from three-phase systems because (1) large amounts of electrical power are most efficiently supplied by three-phase distribution, (2) polyphase rectifiers produce less ripple than do single-phase circuits, and (3) polyphase rectifiers employing gas tubes provide high efficiency and good voltage regulation.

The importance of industrial polyphase rectifiers can be realized from the fact that a surprisingly large portion (perhaps 20 percent) of the total electrical energy developed in the United States passes through such rectifiers. The electrochemical and transportation industries account for a large share of this load, and the remainder represents applications such as in plants where the flexible control of d-c motors justifies the installation of rectifying equipment. Formerly, motor generators and rotary converters provided the conversion from alternating to direct power, but the efficiency, quietness, and serviceability of modern rectifiers has practically outmoded rotary equipment for most of these applications.

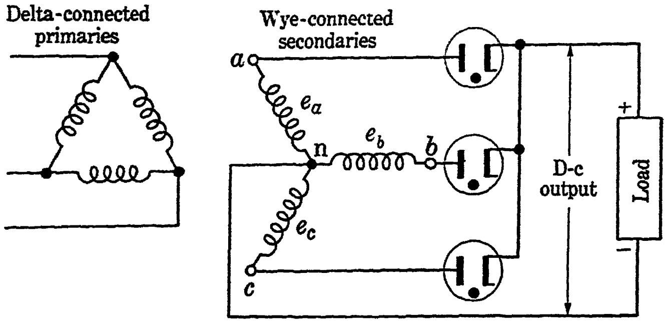
In this chapter we shall first examine several basic rectifier circuits and then discuss the details of filtering, choice of rectifier tubes, etc.

8.1 Three-phase Y Connection

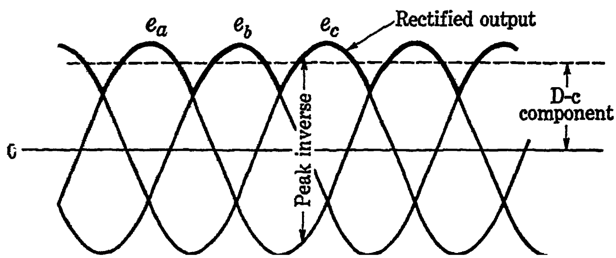
The three-phase Y-connected rectifier of Fig. 8.1 represents a logical extension of the full-wave single-phase rectifier discussed in Chap. 3. Three transformers (or a single three-phase transformer) provide equal three-phase secondary voltages to the three rectifying elements. In accordance with custom, windings of corresponding phase are shown parallel on the diagram, and the angle

of orientation indicates the relative phase position of the voltage developed by each winding. The circuit shows the thermionic gas diodes that would be used for a unit of relatively small power capacity; a larger installation might employ ignitrons.

Because of the unidirectional properties of the diodes, only the



(a) Circuit diagram. Point n is taken as the reference point for all voltages.



(b) Voltage wave forms. These neglect the effects of tube drop and transformer reactance.

FIG. 8.1. Circuit diagram and wave forms for a three-phase Y-connected rectifier.

transformer secondary providing the highest instantaneous voltage can remain connected to the load circuit, and the rectified output voltage follows along the tops of the individual voltage waves, as shown by Fig. 8.1*b*. Actually, because of tube drop, the output voltage follows a path approximately 15 volts below the envelope shown, but this hardly shows on a diagram drawn for an output of several hundred volts. Also, the transformer reactance pre-

vents the instantaneous switching from one secondary to another required to obtain the output illustrated, and there is a brief transition period that slightly affects the wave shape. These details, of importance to a designer, assume little importance in our discussion.

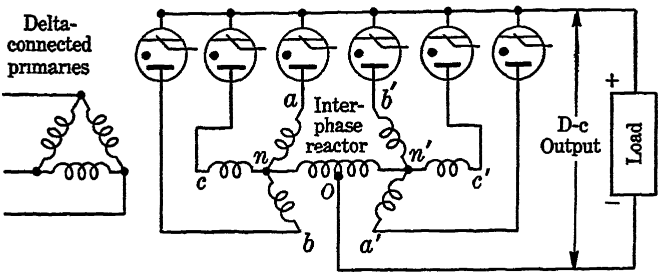
The output wave of Fig. 8.1*b* clearly shows the superiority of the three-phase circuit over the single-phase one in three respects: (1) the average output voltage is higher than for the equivalent single-phase circuit, (2) this higher output is obtained with a lower inverse peak voltage, and (3) the amount of ripple is less and it is of higher frequency. The advantage of higher output voltage is offset by the fact that each secondary and tube operates only one-third of the time. Even so, the effective transformer utilization is somewhat greater than for a single-phase circuit. For a 6- or 12-phase circuit, however, the poor time utilization of a given tube and transformer secondary becomes a disadvantage.

8.2 Double Y Connection

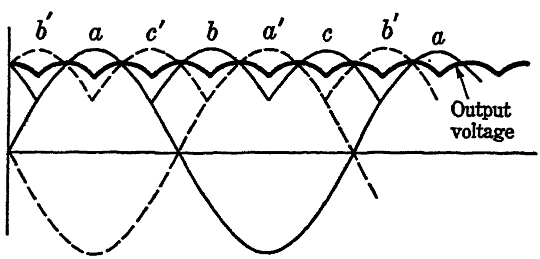
Probably one of the most important polyphase rectifier circuits is the double Y connection, shown by Fig. 8.2. This circuit consists of two parallel-connected three-phase Y's constructed from transformers with twin secondary windings. The Y *abc* utilizes one secondary of each phase, and the second Y *a′b′c′* employs the other three connected in opposite phase. Since the two circuits do not produce the same instantaneous output voltage, the two neutrals n and n' operate at an alternating difference of potential and cannot be directly connected together. This is the reason for the center-tapped interphase transformer shown on the diagram. However, the two identical Y-connected rectifiers do produce the same *average* output voltage, and there is no d-c difference of potential between the ends of the interphase transformer.

The wave forms of Fig. 8.2*b* show the set of three-phase voltages e_a, e_b, and e_c, and the second set e'_a, e'_b, e'_c, which are in phase opposition to the first group. For the purpose of analysis we shall first consider each half of the circuit alone. Thus each Y produces an output voltage with three humps per cycle, exactly like that shown for the Y circuit of Fig. 8.1. The wave forms also plainly show that the two Y circuits produce the same output voltage except with the humps displaced from one another so that one circuit reaches its peak when the other drops to the minimum

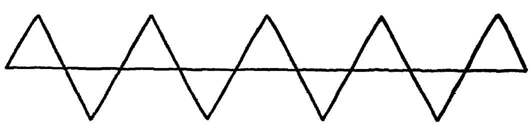
point. Since both circuits have a common positive terminal, this difference must appear between the two neutral points n and n' with an irregular wave form, shown by Fig. 8.2c. The interphase transformer connected to the two neutrals acts as a center-tapped



(a) Circuit diagram. For simplicity no ignitron firing circuits are shown



(b) Wave forms showing the output of each wye and the net output voltage. Wave forms neglect effect of transformer reactance



(c) Voltage $e_{n-n'}$ appearing across the inter-phase reactor

FIG. 8.2. The double Y connection.

transformer with the voltage at center point o always midway between the two ends. Thus the net output voltage between point o and the common positive terminal stays exactly halfway between the two individual circuit output voltages. Figure 8.2b shows this average as a heavy line.

The advantages of this arrangement are fairly obvious. (1) The

combined circuits provide an output with relatively small ripple and with a fundamental ripple frequency equal to six times the incoming line frequency. (2) Although the output is similar to that obtained with a six-phase circuit, each tube and transformer secondary operates a full one-third of the time. (3) The transformer primary utilization is better than that of a simple three-phase Y because each primary has two secondaries and operates two-thirds of the time. This statement is an oversimplification of the situation, but a complete analysis indicates the expected improvement.

8.3 Three-phase Bridge

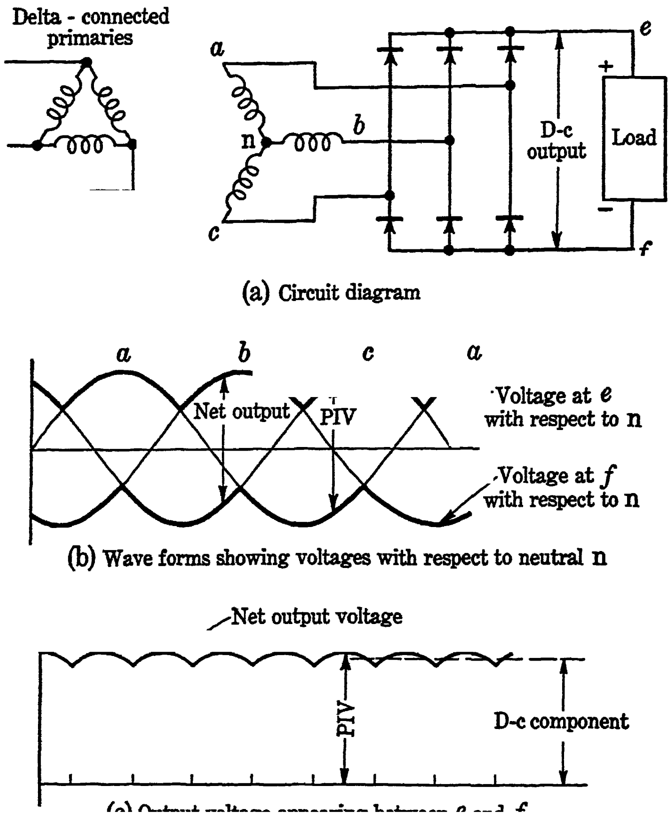
One of the most effective circuits from the standpoint of transformer utilization is the three-phase bridge of Fig. 8.3. While it does not look like any kind of familiar bridge circuit, the circuit arrangement is related to that of the single-phase bridge of Chap. 3. This can be seen by eliminating any vertical pair of rectifying elements which reduces the circuit to that of the single-phase bridge.

The chief disadvantage of the circuit is that only three of the rectifying elements have common cathodes. This rules out the use of a mercury-arc rectifier and complicates an ignitron firing circuit. Even with gas-filled or vacuum diodes the additional filament winding required to heat separate cathodes complicates matters somewhat, although this is not an extremely important consideration. For this reason, the circuit of Fig. 8.3 is shown with contact rectifiers to which it is admirably suited.

A convenient method of analysis is to consider the bridge circuit as composed of two three-phase Y rectifiers. One of these consists of the transformer windings and the top three rectifying elements. These produce a positive voltage at point e, with respect to the neutral n, that follows the positive peaks of the three-phase voltages, as shown in part b of Fig. 8.3. The lower three rectifiers are connected in exactly the same fashion except for being reversed so that they can conduct only on the *negative* peaks of the secondary output voltages. Thus they form a Y-connected rectifier which produces a negative output voltage at point f that follows the negative peaks, as indicated on the same diagram.

Of course we are not particularly interested in the voltage to neutral; the output voltage appearing between e and f is the one of interest. Figure 8.3c shows this output curve obtained from b

by taking the vertical distance between the two heavy curves. The remarkable property of this circuit is that it not only produces a d-c output voltage twice as great as does the ordinary Y connection, but it reduces the ripple to the same relative amount as that



(a) Circuit diagram

(b) Wave forms showing voltages with respect to neutral n

FIG. 8.3. Circuit diagram and wave forms for a three-phase bridge rectifier.

produced by a six-phase circuit. Furthermore, it produces this double output voltage without any increase in peak inverse voltage as compared with the straight three-phase Y. For a three-phase bridge the peak inverse voltage applied to the rectifying elements is only 1.05 times the d-c output voltage. In addition to all this, the transformer utilization is particularly high because each secondary winding operates twice during each cycle.

To counterbalance these advantages there is, of course, the disadvantage regarding the common cathode connections mentioned above and the additional fact that at any moment two rectifying elements in series must conduct to carry the load current. This doubles the effective voltage loss in the rectifiers, which tends to reduce the efficiency.

8.4 Smoothing Filters

Since a polyphase rectifier provides a much smoother output voltage than does a single-phase circuit, there are many applications which require no additional smoothing. For electrochemical applications even the output of a three-phase Y will do, while the considerably smoother output of a double Y circuit is adequate for most industrial applications involving d-c machinery. Consequently, high-power rectifiers seldom employ any filters to reduce the output ripple.

Of course there are applications which require relatively large amounts of well-smoothed output power—a large radio transmitter, for example. In this case the rectifier output is passed through an L-section filter, exactly as discussed in Chap. 3. As compared with a single-phase circuit, a polyphase circuit produces a much smaller amount of ripple; for a three-phase Y the rms fundamental component of ripple is only 17 percent of the direct output voltage, while a circuit producing an output similar to that of the double Y connection reduces this value to only 4 percent. In addition to this, the relatively high ripple frequency makes a given amount of inductance and capacitance especially effective, as shown by Eq. (3.2) for the smoothing factor.

Capacitance input filters are never used for polyphase rectifiers employing gas tubes because of the high peak current and resultant low efficiency. In addition to this, the large capacitor charging current demanded for the first few cycles of operation may cause a thermionic gas diode to exceed the allowable peak current and damage the cathode through positive-ion bombardment.

8.5 Phase-shift Control of Output Voltage

When a rectifier employs thyratrons or ignitrons as the rectifying elements, it becomes possible to control the average output voltage by delaying the firing point at which conduction switches from

one tube to the next. This is illustrated by the wave forms of Fig. 8.4, showing the output voltage for no delay and also for

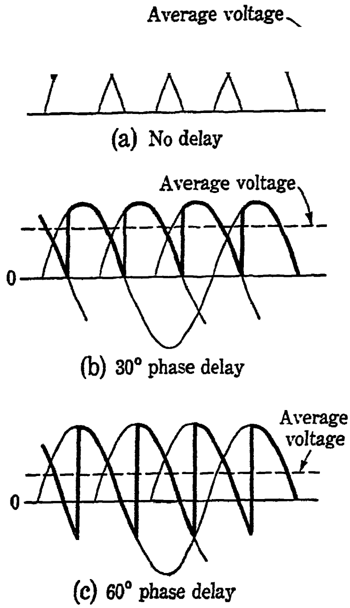


FIG. 8.4. Output-voltage wave shapes showing the effect of phase shift upon the average value.

delays of 30 and 60 degrees. In large power rectifiers the usual purpose of doing this is not to control the output voltage over a wide range, but automatically to adjust the output voltage to keep it constant despite variations of load current or alternating input voltage. For this purpose the necessary phase delay is rather small.

For some applications it is desirable to provide a full range of adjustment which requires a greater delay than 30 degrees and a smoothing inductance in series with the load to prevent the current from dropping to zero during part of the cycle. With this inductance a given tube is forced to continue conduction until the delayed firing of the next one, even though the transformer supply voltage goes temporarily negative.

8.6 Ignitron Firing

A rectifier, such as that of Fig. 8.2 employing ignitrons, requires additional circuits to supply current impulses to the ignitors for

firing the tubes at the correct point in each cycle. Although this entails additional circuit complications, the low arc drop and serviceability of ignitrons makes this complication well worth while. Consequently, ignitron rectifiers have practically displaced multianode mercury-arc rectifiers for voltages below about 1,000 volts. Of course the firing arrangement does waste some energy, but the igniting power is practically independent of the size of the ignitron so that the efficiency of high-power rectifiers remains high. For this reason ignitrons are not ordinarily used in units with power-output ratings of less than about 50 kilowatts.

The firing system employing parallel mercury-vapor diodes, as shown in Fig. 5.19 of Chap. 5, has the serious disadvantage that the firing current must pass through the load. If the load current drops too low, the current may be insufficient to produce reliable firing. For this reason a separate firing circuit independent of the load is better.

There are at present two basic types of firing circuits that have been developed for use in ignitron rectifiers. One type employs thyratrons together with phase control and pulse-forming circuits to deliver a sharp current surge to each ignitor. This arrangement requires one thyratron and its circuit components for each ignitron, and the drawbacks of this system are essentially those of any thyratron: (1) limited life, (2) delay for cathode heating, and (3) easily damaged cathode.

From the standpoint of ruggedness the alternate arrangement of Fig. 8.5 has many advantages. This circuit employs only inductors, capacitors, and copper-oxide or selenium rectifiers, all of which have indefinite life expectancies. Operation is based on the principle that the permeability and thus the inductance of an iron-clad circuit drops to very low values above the saturation point. The saturable reactor with this characteristic is constructed with no air gap, a relatively large number of turns of wire, and a limited cross section of iron. The linear reactor, on the other hand, maintains an approximately constant inductance because of a relatively large air gap and a generous amount of iron.

In this circuit, just as the alternating voltage across C reaches its maximum, the small lagging current through the saturable reactor causes saturation and a sudden decrease in impedance. This permits C to discharge rapidly and produces a sudden current pulse shown by the wave form of i_0 on the diagram. On the reverse

half cycle the capacitor builds up in the opposite polarity, and a pulse flows in the other direction. Since a single ignitron requires only one pulse per cycle, the output of the saturable reactor passes

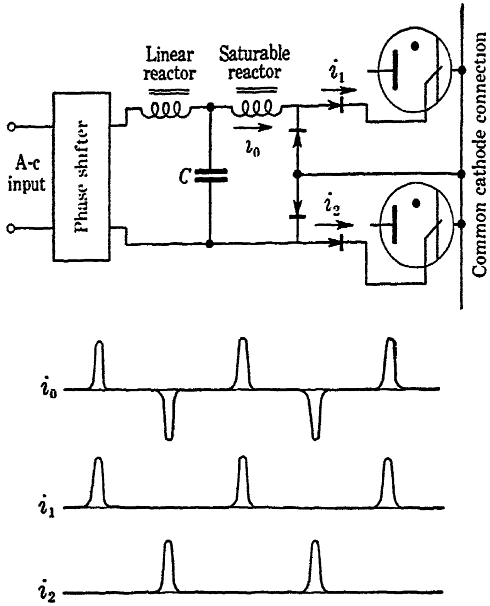


FIG. 8.5. Circuit diagram and wave forms illustrating the operation of a saturable reactor firing circuit for ignitrons.

through a group of disk rectifiers which direct the positive pulses to one ignitron and the rectified negative pulses to the other. Thus the single-pulse-forming circuit can handle the firing requirements of two ignitrons that need to be fired in phase opposition. This arrangement is especially effective with the double Y-connected rectifier of Fig. 8.2, which contains six ignitrons that can be controlled by three such firing circuits with a-c inputs staggered 120 degrees in phase position.

8.7 Efficiency and Voltage Regulation

Since the voltage drop in a gas-filled rectifier is low, and because transformers are exceptionally efficient devices, the over-all efficiency of a large power rectifier unit is very good, especially at the higher voltages where the tube loss becomes a small part of

the total voltage. Figure 8.6, which gives typical efficiency data for ignitron rectifiers of several hundred kilowatt capacity, shows that a rectifier outperforms mechanical equipment, especially at

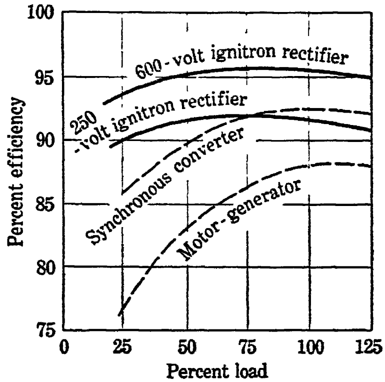


FIG. 8.6. An efficiency comparison of ignitron rectifiers and rotating machinery for the production of direct current.

light loads. At an output of about 100 volts, the 12-volt loss in a gas tube drops the efficiency below that obtainable with rotating machinery and the choice must be based upon other considerations. At very low voltages, however, the three-phase bridge circuit with selenium rectifiers gives good efficiency combined with ruggedness and simplicity. For instance, a 15-volt 100-ampere rectifier of this type may operate at a full-load efficiency of 85 percent.

The natural voltage regulation of a rectifier using mercury-pool tubes is illustrated by Fig. 8.7, which shows that the output voltage

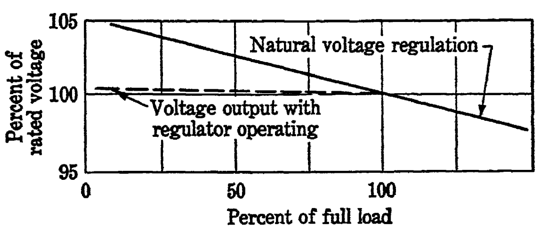


FIG. 8.7. Typical regulation curve for an ignitron rectifier.

drops about 5 percent from zero to full load. This corresponds to the normal regulation curves for mechanical d-c generators so that rectifiers can be operated in parallel with each other or with

existing rotary generating equipment. For situations requiring a more constant output voltage, the phase-shift system can be coupled to the rectifier output to hold the direct voltage automatically at a substantially constant level, as indicated on the curve. This is often done with the aid of saturable reactors in the phase-shifting network; these reactors possess a d-c winding which varies the saturation of the reactor in proportion to the d-c output voltage. This variable reactance in turn shifts the phase of the voltage applied to the firing circuit.

PROBLEMS

8.1 Show that for the idealized wave form of Fig. 8.1 the average output voltage equals 0.827 of the peak alternating voltage.

8.2 In an actual rectifier the gas-filled tube voltage drop subtracts a constant amount from the ideal wave form of Fig. 8.1. If the tube drop is 12 volts, what is the required effective transformer voltage for an average output of 100 volts?

8.3 The output of a double Y-connected rectifier operates into an inductive load that draws a constant current of 200 amp. Compute the peak and average currents carried by any one ignitron.

8.4 For the ideal wave forms of Fig. 8.3 determine the relation between (*a*) the average output voltage and the peak transformer voltage, and (*b*) the average output voltage and the peak output voltage.

8.5 Compute the average output voltage for a Y-connected rectifier with a 60-deg phase delay as compared with the voltage without delay (see Fig. 8.4).

CHAPTER 9

PRACTICAL AMPLIFIER CIRCUITS

THE FUNCTION of an amplifier is to amplify an electrical signal consisting of a simple or a complex voltage wave. An electromechanical transducer such as a microphone or a phonograph pickup may provide this voltage. Sometimes the signal comes from a photoelectric cell; sometimes an antenna supplies a few microvolts of radio frequency for amplification. As a general rule, the signal voltage is small and it is a complex wave containing many frequency components. Figure 9.1 shows the approximate

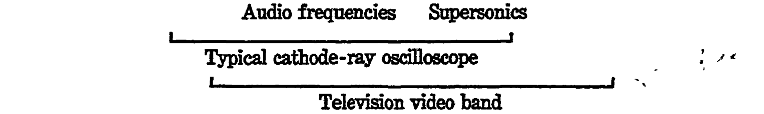


FIG. 9.1. A chart showing the frequency spectrum from 0.1 to 10^8 cycles per second.

frequency bands covered by the various types of signals ranging from very low structural vibrations to the high radio frequencies used for the transmission of frequency modulation. This does not show the whole useful spectrum because the practical frequency range now extends to about 100,000 megacycles. Above 100 megacycles the techniques differ considerably from those employed at lower frequencies, and we shall concern ourselves with frequencies below this limit.

9.1 Amplifier Classification

Figure 9.2 shows an elementary classification of amplifier types. This classification first divides the field into wide-band and tuned

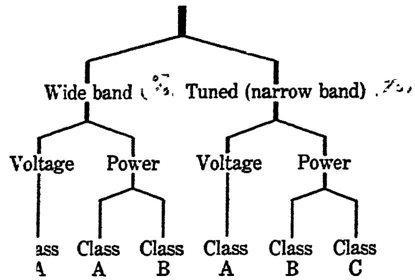


FIG 9.2. An elementary classification of amplifier types.

amplifiers. The term wide-band amplifier indicates one which amplifies a relatively wide band of frequencies equally well. An audio amplifier is a good example of this; to amplify speech and music requires a circuit that amplifies uniformly all frequency components between 50 and 10,000 cycles per second. Figure 9.1 shows that a cathode-ray oscilloscope requires an even better amplifier capable of amplifying all frequencies between perhaps 10 and 200,000 cycles per second. Television needs an extremely wide band because a picture signal contains frequency components ranging from 30 cycles to 4 megacycles per second.

Tuned or narrow-band amplifiers are employed for amplifying a single frequency or a narrow band of frequencies. The tuned amplifiers in radio receivers and transmitters provide examples of this type of service. The classification of wide or narrow band is not based on the absolute band width but rather upon the percentage band width. For example, an audio amplifier handles only a 10-kilocycle band, but the ratio of maximum to minimum frequency is 200. On the other hand, a tuned amplifier capable of handling frequencies between 10.0 and 10.1 megacycles is a narrow-band circuit although the absolute band width exceeds the audio band by a factor of 10.

Wide- and narrow-band amplifiers may be designed as either voltage or power amplifiers. Voltage amplifiers function to provide the maximum voltage gain and commonly serve to provide

an amplified signal for the following stage of amplification. Power amplifiers are needed where the load demands power from the amplifier. A loudspeaker, for example, cannot operate on voltage alone. Actual power must be delivered to it for conversion into sound energy and heat losses. This requires the amplifier to provide a substantial output current as well as voltage.

The diagram of Fig. 9.2 indicates a further subdivision of amplifiers into classes A, B, and C, as defined by Fig. 9.3. All voltage amplifiers and many small power amplifiers operate in class A, which is characterized by relatively low distortion and low efficiency. The desirability of class B and class C operation is difficult to understand at this point in the discussion because the plate-current waves shown by Fig. 9.3 are obviously badly distorted.

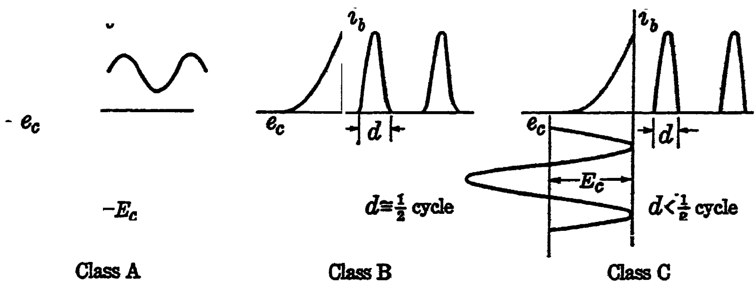


FIG. 9.3. A graphical definition of class A, class B, and class C amplification.

With tuned amplifiers, however, this distortion may be of little consequence because the circuit does not reproduce the harmonics. Class B or C operation has the important advantage of high efficiency—up to 80 percent as compared with 20 percent for class A.

This chapter is devoted to the study of simple wide-band voltage and power amplifiers capable of handling frequency components up to several hundred kilocycles per second.

9.2 Distortion

The term amplifier distortion indicates any departure of the output from being an accurate enlargement of the input wave. In terms of the Fourier concept of representing a complex wave by its harmonic series, distortion can be classified into three types: amplitude distortion, frequency distortion, and phase distortion.

Amplitude Distortion. Figure 9.4 shows an example of amplitude or nonlinear distortion caused by curvature of the vacuum-tube

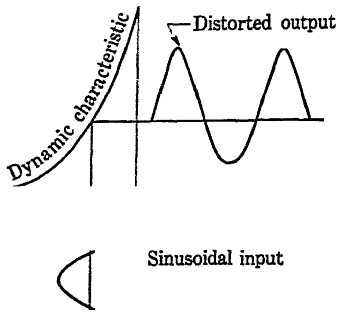


FIG. 9.4. An illustration of amplitude or nonlinear distortion caused by a curved characteristic.

characteristics. Amplitude distortion occurs because the output is not proportional to the input even though a sine wave of single frequency is applied. Resistors, capacitors, and inductors do not normally cause amplitude distortion, but any nonlinear circuit element, usually the vacuum tubes, may produce this distortion if the signal voltage is large.

Frequency Distortion. Frequency distortion occurs when an amplifier does not amplify equally well all frequencies *within a desired band.* This distortion can be observed by measuring the amplification of a circuit over a wide band of frequencies and plotting a curve of amplification against frequency, as shown by Fig. 9.5. The curve shown exhibits no frequency distortion between 100

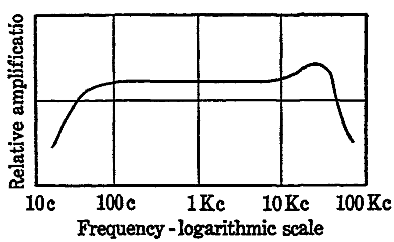


FIG. 9.5. Typical frequency-response curve of an audio amplifier. The curve exhibits negligible frequency distortion between 100 cycles and 10 kilocycles.

and 10,000 cycles per second, and the amplifier would be suitable for a public-address system.

Figure 9.6 shows the effect of frequency distortion on a complex wave consisting of a fundamental and a second harmonic only.

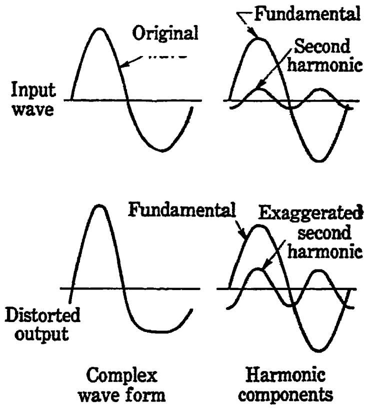


FIG. 9.6. This shows the effect of frequency distortion on a complex wave in which the second harmonic is amplified more than the fundamental.

The circuit amplifies the second harmonic more than the fundamental (as an example), and the output wave looks different from the input. It is easy to confuse the resulting distortion with the amplitude distortion depicted by Fig. 9.4.

Phase Distortion. Phase distortion occurs when an amplifier shifts the relative phase positions of the various signal components. Figure 9.7 illustrates this effect. Both the fundamental and the second harmonic have the correct relative amplitudes (no fre-

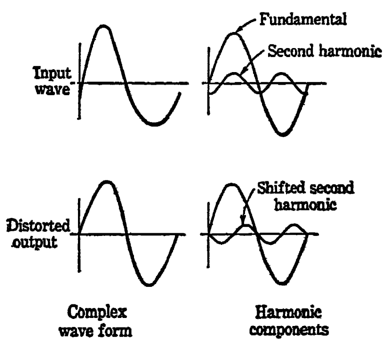


FIG. 9.7. This shows how phase distortion may change the appearance of a wave by shifting the relative positions of the fundamental and the harmonics.

quency distortion), but the phase shift changes the appearance of the output wave. Phase distortion is unimportant in the amplification of speech and music because the ear does not recognize the phase positions of the various components, but the amplifier for a cathode-ray oscillograph must produce very little phase distortion.

Phase and frequency distortion are produced by the capacitors and inductors of an amplifier circuit rather than by the vacuum tube itself. The two types of distortion always occur together because the reactances that produce frequency distortion also cause phase shifts, and an irregular frequency response curve invariably indicates phase distortion.

9.3 The Resistance-capacitance-coupled Voltage Amplifier

The amplifier circuits discussed in Chap. 4 are *single-stage* amplifiers consisting of a single tube and load impedance, together with the associated sources of direct voltage. A practical amplifier must consist of several such stages of amplification in *cascade* with the output of one stage coupled to the next stage, and so on, until sufficient amplification is obtained. All the amplifier stages up to the last one are normally *voltage* amplifiers designed to provide an amplified voltage to drive the grid of the following vacuum tube. The final stage may be either a voltage or a power amplifier, depending upon the characteristics of the device to be operated. In a radio receiver, for example, a final power-amplifier stage provides the energy required to operate a loudspeaker.

One practical solution to the problem of coupling the output of one amplifier stage to the grid of the next one is the resistance-capacitance-coupled amplifier circuit of Fig. 9.8. This useful and

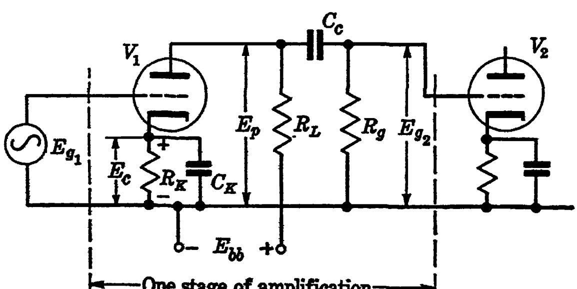


FIG. 9.8. One stage of resistance-capacitance-coupled amplification using a triode.

versatile circuit can be designed to provide wide-band amplification covering the range from less than one cycle per second up to several hundred kilocycles. The circuit is essentially an adaptation of the basic amplifier of Fig. 4.8 with improvements to eliminate the need for a bias battery and to apply the alternating voltage developed across R_L to the grid of the second tube.

The heavy line across the bottom of the diagram represents the *common*, or "ground," lead to which a large number of the circuit connections are made. The metal chassis supporting the circuit components usually serves as this connection. Resistor R_k and capacitor C_k eliminate the need for a separate grid-bias battery, as suggested by the equivalent circuits of Fig. 9.9. Raising the

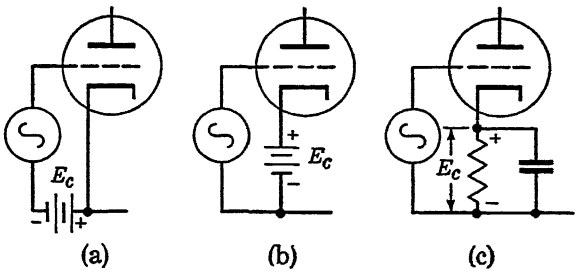


FIG. 9.9. Alternate methods of obtaining the grid-bias voltage. The self-bias arrangement of (*c*) eliminates the need for an extra battery.

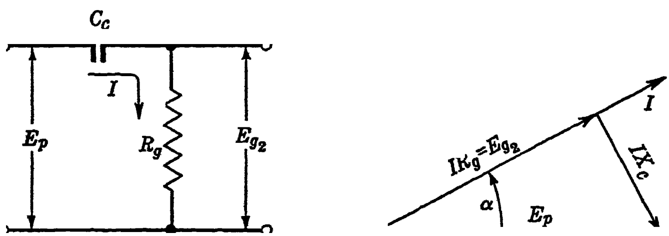
cathode potential has the same relative effect as depressing the grid voltage, and this arrangement eliminates the need for an additional source of voltage. The value of R_k is easily computed from Ohm's law; for example, an 8-volt bias for a tube carrying a plate current I_b of 4 milliamperes requires a cathode-bias resistor of 2,000 ohms. Without C_k the bias developed by R_k is not steady because the alternating component of plate current I_p produces alternations in the voltage drop. The function of C_k is to "by-pass" the alternating components of current. With C_k the voltage drop across the parallel combination consists of a d-c component I_bR_k plus an a-c component nearly equal to I_p times the reactance of C_k. With a sufficiently large condenser to reduce this reactance, the a-c ripple across the bias network can be made negligible. A capacitance of 10 microfarads or more provides reasonably smooth bias for an audio amplifier. Since X_c is inversely proportional to the frequency, low-frequency amplifiers may require an impossibly large by-pass condenser. Omitting C_k

does no particular harm; the alternating voltage developed across R_k reduces the net alternating grid-cathode voltage and decreases the effective amplification by a factor of two or more.

Resistor R_L of Fig. 9.8 serves as the load resistor across which the vacuum tube develops an amplified output voltage for the following tube. Coupling capacitor C_c serves to pass the alternating component of voltage to the grid of the second tube while simultaneously blocking off the direct component of voltage which would upset the bias on the second grid. To do this C_c must present a negligible series reactance to the signal frequencies. This is easily accomplished with a reasonable size of capacitance.

Resistor R_g, commonly called the grid leak, serves to maintain the correct bias on the second tube. Circuit elements are seldom ideal. The insulation of C_c is not perfect, nor is the grid current of V_2 absolutely zero. With this in mind let us imagine that the input signal is temporarily removed so that no alternating components of voltage occur in the circuit. Under this condition the voltage on the grid of V_2 should be zero with respect to the common chassis connection because the cathode resistor provides the grid bias. Resistor R_g provides a path for the tiny condenser leakage current and grid current so that they produce a negligible voltage drop and affect the bias voltage an inappreciable amount. Fortunately the leakage currents are so minute (around 10^{-9} ampere) that R_g may be as high as a megohm or more. The vacuum-tube design handbooks commonly give the allowable upper limit for R_g as part of the design information for each tube.

With an alternating signal reapplied to the circuit, the tube V_1 develops an alternating voltage E_p across R_L. As suggested by the partial circuit of Fig. 9.10, this alternating voltage appears across C_c and R_g in series. In general, E_p exceeds E_{g_2}, as shown



(a) Circuit diagram (b) Vector diagram

FIG. 9.10. An analysis of the effect of the C_c and R_g on the low-frequency response of an R–C amplifier.

by the vector diagram, but with X_c small compared with R_g the coupling network (C_c, R_g) introduces practically no loss. This can never be true at all frequencies because X_c is inversely proportional to the frequency, and at some low frequency X_c equals R_g. Below this frequency the output voltage E_{g_2} and the net circuit amplification rapidly decrease.

Frequency Response of an R–C Amplifier. Figure 9.11 shows a

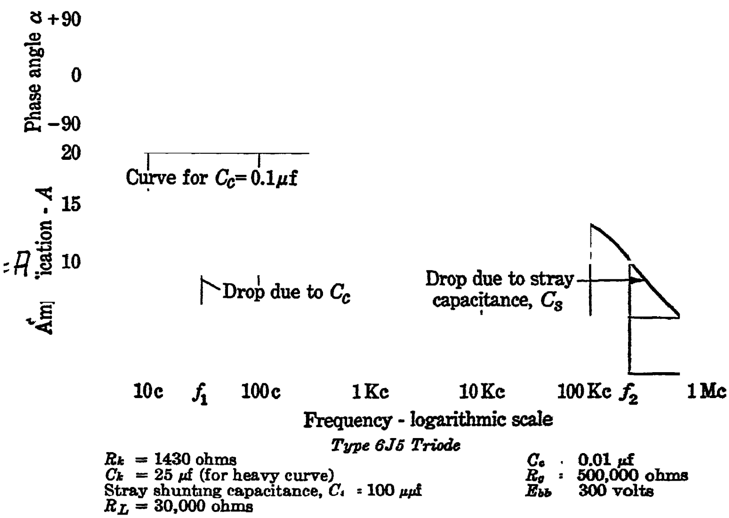


FIG. 9.11. Frequency-response and phase-shift curves for the single-stage R–C amplifier of Fig. 9.8.

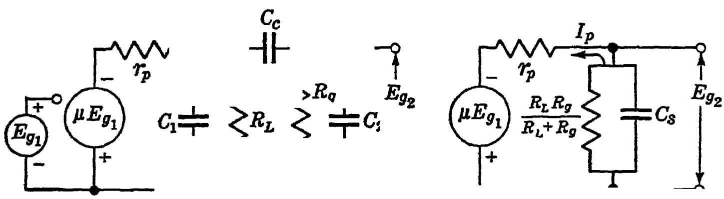
curve of amplification versus frequency for the circuit of Fig. 9.8 with representative values of circuit components. Over several decades the curve is flat, which indicates that the capacitive reactances in the circuit have a negligible effect and that the amplification is determined only by the resistances and the choice of tube. This flat part is called the mid-frequency region. At the low-frequency end the curve drops off because of coupling capacitor C_c, as discussed in the preceding paragraph.

The point at which the amplification drops to $1/\sqrt{2}$ times the mid-frequency value is arbitrarily called the low-frequency cutoff point, marked f_1 on the diagram. Except for a small correction due to the cathode bias circuit and the effect of R_g on the load resistance, this point occurs at a frequency for which X_c equals R_g.

This makes the vector diagram of Fig. 9.10 a 45-degree triangle with voltage E_{g_2} equal to $0.707E_p$. There is also a leading phase shift of 45 degrees at this frequency. The size of C_c determines the low-frequency cutoff point; increasing C_c ten times shifts the cutoff point to one-tenth of the frequency, as suggested by the dotted curve of Fig. 9.11.

At the high-frequency end of the curve the amplification also drops, not because of any capacitance intentionally placed in the circuit, but because of the internal tube capacitances and the inevitable stray shunting capacitance between the wiring and the metal chassis. This total capacitance C_s is difficult to keep low; the value of 100 micromicrofarads given by Fig. 9.11 represents an average value for a triode circuit.

The equivalent circuits of Fig. 9.12 help to show more clearly



(a) Complete equivalent circuit (b) High-frequency equivalent

FIG. 9.12. Simplified equivalent circuits for an R–C amplifier.

the effect of the stray shunting capacitance on the amplifier performance. Figure 9.12*a* represents the complete equivalent circuit with the tube replaced by an equivalent generator and its plate resistance r_p. Capacitance C_1 consists of the plate-cathode capacitance of the tube plus the tube socket and wiring connections to the left of C. Capacitance C_2 represents the wiring, socket, and effective grid-cathode capacitance of the second tube. Figure 9.12*b* shows the simplified circuit for the high frequencies. Capacitor

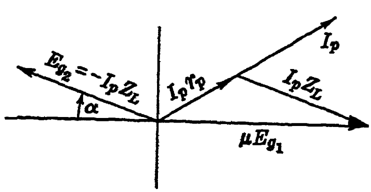


FIG. 9.13. Vector diagram showing the voltages and currents of Fig. 9.12*a*. At high frequencies the capacitive load impedance makes I_p lead μE_g, but the output voltage E_{g_2} *lags* with respect to the normal 180-degree phase shift.

C_c has an appreciable reactance at low frequencies only; at the high frequencies it becomes negligible and is omitted from the diagram. This places R_L and R_g in parallel, and they can be replaced by a single equivalent resistor. The diagram also shows C_1 and C_2 combined into single capacitor C_s. This simplified circuit shows plainly that the equivalent vacuum-tube load impedance consists of a parallel resistance and reactance. At mid frequencies X_s is so large compared with the load resistance that the amplification is unaffected, but at high frequencies the reactance drops, the load impedance decreases, and the amplification becomes less. A lagging phase shift accompanies this drop, as analyzed by the vector diagram of Fig. 9.13.

Pentode R–C Amplifier. With a few modifications to provide screen voltage the resistance-capacitance-coupled amplifier works equally well with pentodes. Figure 9.14 shows one circuit change

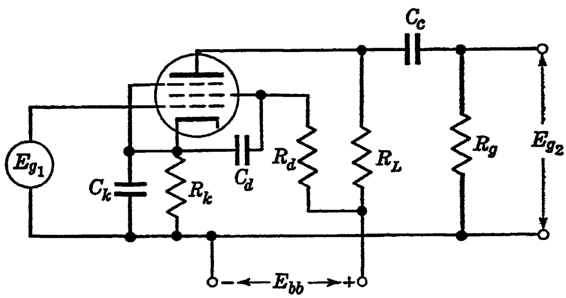


FIG. 9.14. A basic single-stage pentode R–C amplifier.

to consist of the addition of resistor R_d to drop the supply voltage to a value suitable for the screen, usually between 50 and 100 volts. The screen current flowing through R_d produces the required drop. For instance, with a supply voltage of 300 volts, a desired screen voltage of 100 volts, and a screen current of 0.2 milliampere, the necessary screen dropping resistor is 200/0.0002, or 1 megohm. With an applied signal the screen current also contains alternating components, and condenser C_d provides a low-reactance path to keep them out of R_d. This stabilizes the screen voltage in exactly the same fashion as C_k smooths the bias voltage developed across R_k.

As explained in Sec. 4.11, pentodes give more amplification than do triodes and a higher load resistance is worth while. In the circuit of Fig. 9.14 the load resistance is commonly larger than

100,000 ohms. For this reason and also because the plate resistance is high, the effect of the stray capacitance is greater than in a triode circuit with its relatively low resistances. Consequently, an ordinary high-gain pentode R-C amplifier usually has a poorer high-frequency response than that shown by the curve of Fig. 9.11.

Extended High-frequency Response. Television video amplifiers and the better types of wide-band cathode-ray oscilloscope amplifiers need to provide flat response over a very wide frequency range. The low-frequency cutoff can be adjusted by the choice of coupling condenser, but the high-frequency response is controlled by the care taken in reducing C_s to a minimum. With pentodes, which have lower input capacitance than do triodes, and careful wiring, the shunt capacitance can be reduced to less than 30 micromicrofarads. Further extension of the high-frequency cutoff must be made at the sacrifice of amplification by reducing the load resistance. To this end, special pentodes are available which possess a very high transconductance. A type 6AC7 television pentode, for example, has a transconductance of about 10,000 micromhos. With a load resistance of only 3,000 ohms the gain provided by this tube is

$$A = -g_mR_L = -(0.01)(3{,}000) = -30$$

This is better than the normal triode amplifier of Fig. 9.8.

The high-frequency drop in gain becomes serious at the point where X_s equals R_L. Equating the two to find the corresponding frequency, we obtain

$$X_s = R_L = \frac{1}{2\pi f_2 C_s}$$

$$f_2 = \frac{1}{2\pi R_L C_s} \tag{9.1}$$

This gives a cutoff frequency of 1.8 megacycles for a shunt capacitance of 30 micromicrofarads and a load resistance of 3,000 ohms. Further extensions in high-frequency response require the use of complex coupling networks and correspondingly complex theoretical considerations.

9.4 The Direct-coupled Amplifier

The direct-coupled amplifier of Fig. 9.15 eliminates the coupling condenser between stages and thus produces a flat response curve

extending to zero frequency. This type of circuit is used for laboratory work to amplify slowly varying signals and waves with long constant periods such as those experienced in slow structural

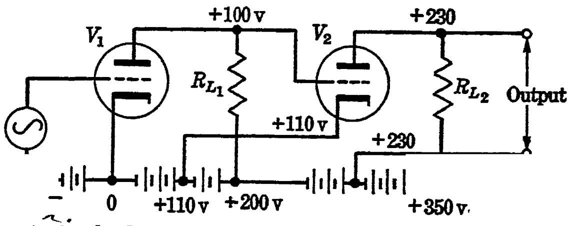


FIG. 9.15. A simple direct-coupled amplifier which eliminates the coupling capacitor C_c and thus removes the low-frequency limitation on the frequency-response curve.

vibrations and in medical electroencephalographic work. In another type of application, a direct-coupled amplifier in combination with a portable meter to measure the amplified changes in a tiny voltage or current can often replace a delicate laboratory-type meter.

The circuit details of Fig. 9.15 are simple. The plate of the first amplifier tube is directly coupled to the grid of the second. This places the grid at a positive potential with respect to the zero point shown on the diagram. To provide a negative grid bias for the second tube the cathode must exceed the grid in potential. Connecting the cathode to the plus 110-volt point accomplishes this and provides 10 volts of bias for tube V_2. The same arrangement can be repeated for additional stages.

Despite its simplicity, this type of amplifier suffers from two serious defects which limit its use to applications absolutely requiring extended low-frequency response. These defects are as follows:

1. The d-c supply voltage is relatively high because the supply voltages for each stage are effectively connected in series. In addition, the power supply must provide a large number of voltage taps, each accurately held at the proper voltage. If, for instance, the 110-volt tap should change by 1 percent without a compensating change in the other potentials, the bias on V_2 would change 10 percent. In the R-C amplifier, on the other hand, a single supply voltage will take care of a number of stages.

2. For an amplifier with more than two stages, drift becomes

serious. By drift is meant changes in the operating point caused by supply-voltage variations, changes in cathode temperature, thermal expansion of the tube elements, or a variety of other causes. A 10-millivolt change in bias voltage E_{c_1}, for example, after being amplified 30 times by V_1 and 30 times by V_2 produces a 9-volt shift in the plate voltage of V_2. This would be enough to move the bias of a third tube completely out of the proper linear range of operation. Likewise, a change of cathode temperature at V_1 changes the emission velocity of the electrons and produces a small plate-current change equivalent in effect to a drift of voltage E_{c_1} with a corresponding effect upon the operating point.

Balanced or bridge circuits have been devised which cancel the majority of the drift caused by d-c supply voltage and cathode-voltage variations and which can maintain even a three-stage amplifier at approximately the correct operating point. The residual drift, however, still creates an ambiguity as to whether the slow changes in the output represent amplifier instability or actual changes in the input signal.

9.5 The Transformer-coupled Amplifier

The transformer-coupled amplifier of Fig. 9.16 has two advantages compared with the R–C amplifier: (1) the transformer

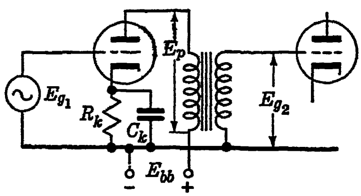


FIG. 9.16. A transformer-coupled voltage amplifier.

provides a voltage step up in addition to that produced by the tube, and (2) the low d-c resistance of the transformer primary (plate) winding produces an almost negligible d-c voltage drop. Advantage 1 is no longer of much importance. Interstage transformers can be used only with triodes, but an R–C coupled pentode amplifier provides more gain than a transformer-coupled triode. Advantage 2 is sometimes important. The saving of 100 volts or more for the plate-voltage supply is certainly important for a portable battery amplifier, but the economy is less valuable in an a-c operated system.

From the cost standpoint a transformer is heavier, larger, and more expensive than the two resistors and single capacitor required for an R–C amplifier. Furthermore, an R-C amplifier provides better frequency and phase response. Figure 9.17 shows

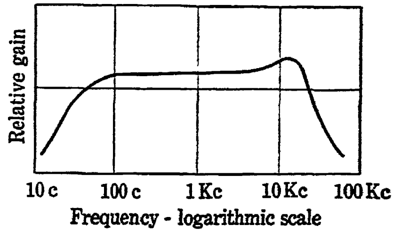


FIG. 9.17. Typical frequency characteristics for a transformer-coupled voltage amplifier.

a typical response curve for a good transformer-coupled circuit. The curve remains flat over the full audio-frequency band, but it is impossible to extend the range to cover the much wider band possible with resistance-capacitance coupling.

A transformer represents a rather complex load impedance, and the flat part of the amplification curve is due, not to the uniformity of the load impedance, but to the fact that the amplification of a triode remains nearly independent of the load as long as the load impedance greatly exceeds the plate resistance. Figure 4.25 illustrates this fact, and Eq. (4.17) also shows this to be true.

$$A = \frac{-\mu Z_L}{r_p + Z_L} \tag{4.17}$$

With Z_L much larger than r_p the gain approaches the amplification factor. The amplification computed from this equation is based on the alternating plate voltage developed across the transformer primary. However, the secondary voltage is larger than the primary by the turns ratio, N. Thus the mid-frequency gain of a transformer-coupled stage approaches

$$\boxed{A = \mu N} \tag{9.2}$$

At low frequencies the inductive reactance of the transformer primary decreases to the point where r_p exceeds Z_L and the amplification drops. To ensure good low-frequency response an audio transformer must have a high-permeability alloy core wound with many turns of wire. Unfortunately, the large number of

turns increases the winding capacitance which adversely affects the high-frequency response.

Transformer coupling with a pentode gives very poor response. The high pentode plate resistance makes the gain depend almost entirely upon the value of the load impedance.

$$A = -g_m Z_L$$

Thus, irregularities in the transformer impedance vs. frequency curve affect the amplifier gain and the response curve no longer remains flat over the mid-frequency range.

The high-frequency response is controlled by the transformer winding capacitance and the leakage inductance of the windings. Any attempt to increase the amplification by increasing the number of secondary turns also increases this capacitance and lowers the point of high-frequency cutoff. For this reason a limiting turns ratio of about three has been found for transformers covering the audio-frequency band.

Transformer coupling must be employed for a circuit in which the grid of the second tube is driven positive during part of each cycle. The grid current produced can pass through the transformer secondary winding without appreciably changing the grid bias, but with resistance coupling the voltage drop in R_g seriously affects the bias.

9.6 Impedance Transformation

The voltage transformation provided by a transformer is well known to most engineers, but the properties of current and impedance transformation are seldom common knowledge. Since impedance transformation or "matching" occurs in many electronic circuits, we shall briefly study the transformer from this standpoint.

Figure 9.18 shows a transformer connected to a load impedance.

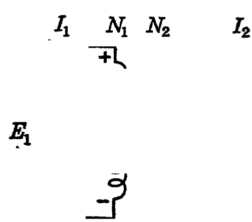


FIG. 9.18. Diagram showing a transformer coupled to a load impedance.

Transformers are efficient devices, and their operation approaches that of an idealized concept called the ideal transformer. This ideal transformer has no power losses (a practical one may be 95 percent efficient), the exciting current is negligibly small (this amounts to using iron of infinite permeability), and all the magnetic flux completely links both coils (this neglects the leakage flux). The following conclusions, drawn for an ideal transformer, never exactly apply to a practical one, but the difference between the two is reasonably small.

In the ideal transformer of Fig. 9.18 the alternating flux produced by the primary winding links both windings completely. Therefore the voltage induced in each turn is the same, and the total voltages are proportional to the numbers of turns. Therefore,

$$\frac{E_1}{E_2} = \frac{N_1}{N_2} \tag{9.3}$$

With Z_2 disconnected, no secondary current flows and I_1 drops to zero (at least very small for a practical transformer). There must still be an alternating flux in the core, however, to generate a counter-emf in the primary that just opposes the applied voltage. With Z_2 connected, current I_2 immediately flows and I_1 appears in the primary. In fact the product I_1N_1 must exactly equal $-I_2N_2$ to keep the net core flux the same. This flux must stay the same to generate the same primary counter-emf equal to E_1. The minus sign indicates that the direction of the secondary ampere turns must oppose the primary ampere turns. In Fig. 9.18 both windings circle the core in the same direction to give the polarities shown. However, the primary current goes down the primary while I_2 travels up, which reverses its direction around the core. Taking the direction of I_2 shown as the positive direction and considering the minus sign associated with the winding, we have

$$\frac{I_1}{I_2} = \frac{N_2}{N_1} \tag{9.4}$$

This must be true at every instant so that I_1 and I_2 go through zero simultaneously and are in phase.

The apparent impedance looking into the transformer primary is

$$Z_1 = \frac{E_1}{I_1} \tag{9.5}$$

Substituting from Eqs. (9.2) and (9.3) for I_1 and E_1, we get

$$Z_1 = \left(\frac{N_1}{N_2}\right)^2 \frac{E_2}{I_2} = \left(\frac{N_1}{N_2}\right)^2 Z_2 \tag{9.6}$$

Therefore a transformer transforms impedances in proportion to the *square* of its turns ratio.

A given practical transformer can behave in an approximately ideal manner over a restricted range of frequencies, depending upon the excellence of its design. In general, a transformer connected to a load impedance performs better than an interstage coupling transformer without any load. Commercial output transformers can be obtained to cover a range from 20 cycles to 20 kilocycles.

9.7 The Class A Triode Power Amplifier

An amplifier must not only increase the size of the input signal to a reasonable size, but it must also provide sufficient power to operate some device that transforms the signal into a useful form. In a radio receiver the output operates a loudspeaker; a vibration amplifier terminates in an oscillograph for recording the wave forms on photographic film; a radio transmitter supplies power to be radiated by the antenna. An amplifier stage designed to provide this power is called a power amplifier. In its design, voltage amplification is secondary because plenty of amplification can be provided ahead of the power stage. Rather, the circuit is designed for the greatest *product* of alternating output *voltage* and *current*.

Figure 9.19 shows the basic circuit diagram of a triode power

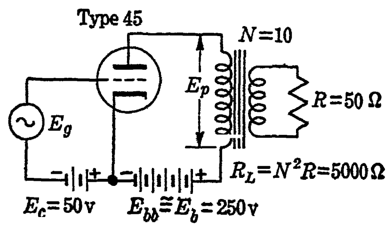


FIG. 9.19. Basic circuit diagram for a triode class A power amplifier.

amplifier. The circuit is already familiar except for having the load resistance connected to the amplifier through a transformer. The transformer serves two functions: First, it transforms resistance R into the desired load resistance for the tube. For example, a transformer to change a 5-ohm loudspeaker into a

2,000-ohm load resistance requires a turns ratio of 20. This makes it possible to design the loudspeaker for best performance without regard to its impedance; the transformer then transforms this into the most desirable load for the vacuum tube. The second function of the transformer is to eliminate the voltage drop that would occur if an actual load resistance were placed in the plate circuit. This improves the efficiency by decreasing the required supply voltage. This improvement occurs because the transformer presents an impedance only to alternating components of current within the frequency range for which it is designed. The d-c component of current I_b passes through the primary without producing any voltage drop other than the small IR drop in the winding resistance.

A power-amplifier tube must not only produce a large alternating output voltage, but it must also provide a large alternating component of plate current. One tube of this type is the type 45 power-amplifier triode, the characteristic curves for which are shown in Fig. 9.20. This tube operates at voltages equivalent

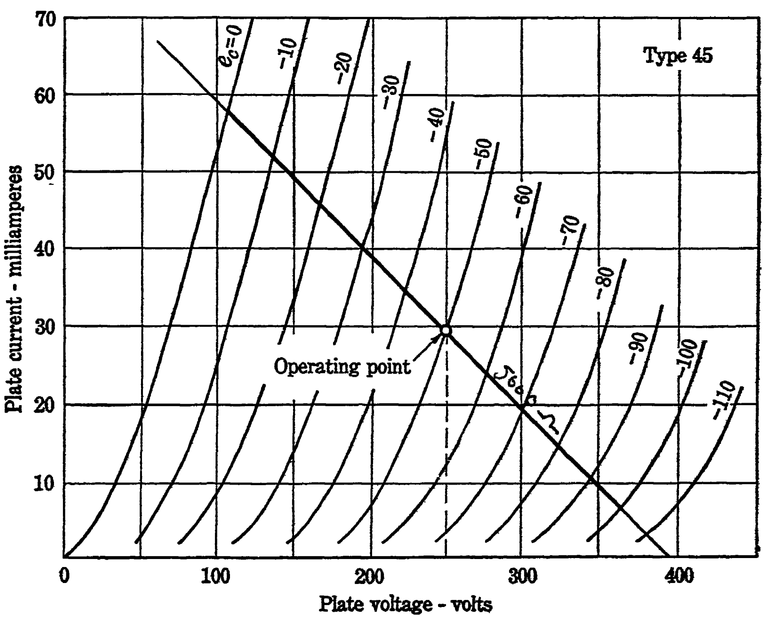


FIG. 9.20. Characteristic curves and load line for a type 45 triode in the power amplifier of Fig. 9.19.

to the voltage-amplifier triodes previously studied, but the current-handling capacity is about ten times as great. The amplification factor is low—only 3.5—but this is a secondary consideration.

Construction of Load Line. The operating point for the amplifier of Fig. 9.19 is determined by the supply voltages E_{bb} and E_c. Neglecting the transformer winding resistance, we assume the actual plate voltage E_b to be 250 volts, and the intersection of the -50 grid-voltage curve with the vertical 250-volt line shows the operating plate current I_{b_0} to be 29 milliamperes.

With an alternating signal voltage applied to the grid, the plate current varies above and below the operating-point value, but to these alternations the transformer presents a resistance of 5,000 ohms. Therefore, the relationship between plate-voltage *changes* and plate-current *changes* is the same as though a 5,000-ohm resistor were actually in the circuit. Thus the point of operation follows along a straight line having a slope of 1/5,000, exactly as discussed in Chap. 4. Figure 9.20 shows this load line. The intersection of the load line with the zero axis at the 395-volt point shows the plate-supply voltage required for a 5,000-ohm load without an intervening transformer. The transformer prevents a voltage loss of 145 volts in this particular case.

To obtain the greatest power output capacity with an input signal that drives the grid just to zero, the load resistance should approximate twice the plate resistance of the triode. This factor of two can be deduced from simple theoretical considerations, but the exact value must be determined from a number of trial-and-error plots on the characteristic curves. It is not difficult to see that a very high load resistance will produce little power output because a level load line on the chart represents large voltage swings with correspondingly tiny current swings; the product of alternating voltage and current is small. Likewise, a very low load resistance (nearly vertical line on the chart) produces large current swings, but the alternating voltage developed is small and again the product is low. At the intermediate point where the load equals twice r_p, the greatest output is obtained with a reasonable distortion.

Computation of Output Power and Efficiency. To illustrate the capabilities of a power amplifier we shall carry out an approximate computation of the maximum power output and efficiency for the load line of Fig. 9.20. This load line has already been adjusted

for correct bias so that an alternating input signal that just swings the grid voltage to zero will not swing it too far negative into the cutoff region. This bias is critical. Figure 9.21 shows 5,000-ohm

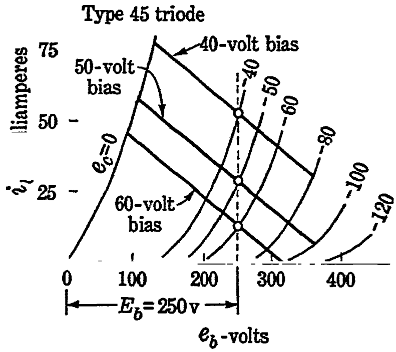


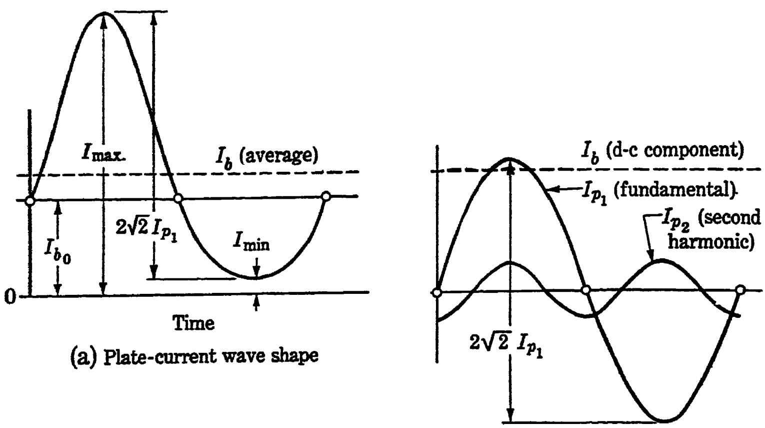
FIG. 9.21. Load lines for several different conditions to show the effect of the grid bias on the distortion and power output of a power amplifier.

load lines for biases of −40, −50, and −60 volts. With a bias of −40 volts, the distortion is small, but the plate current is high and the power output is reduced by the limited grid voltage swing. This produces poor efficiency. With a −60-volt bias the plate current is low, but the distortion is excessive unless the grid voltage swing is held within the limits of the 40- and 80-volt lines. This again reduces the power output and capacity. The 50-volt bias represents the compromise between good efficiency and a reasonable amount of distortion.

For a sinusoidal signal having a maximum value of 50 volts the distortion produced with the load line of Fig. 9.20 is about 6 percent. This means that the effective value of the total harmonic content in the distorted wave is 6 percent of the fundamental component. Most communications and electronics handbooks describe systems for computing the distortion from voltage and current readings taken from the load line.

Figure 9.22 shows the wave form of the distorted plate current. This diagram exaggerates the amount of distortion about three times. The distortion is essentially second harmonic, and Fig. 9.22*b* shows the separate components of the complex wave. This diagram shows two important things: (1) The average plate current with a signal is more than the operating point plate current I_b of 29 milliamperes because the distortion makes the upper half

cycle of the wave larger than the lower half. (2) The difference between the maximum and minimum current values just equals



(a) Plate-current wave shape

(b) Harmonic components

FIG. 9.22. Plate-current wave shapes for the triode amplifier of Figs. 9.19 and 9.20 operating with an alternating grid voltage that just swings the grid to zero.

twice the peak value of the fundamental. Thus, from the diagram,

$$\begin{aligned} I_{max} &= I_b + \sqrt{2}I_{p_1} + \sqrt{2}I_{p_2} \\ I_{min} &= I_b - \sqrt{2}I_{p_1} + \sqrt{2}I_{p_2} \\ I_{max} - I_{min} &= 2\sqrt{2}I_{p_1} \end{aligned} \qquad (9.7)$$

For the particular amplifier of Fig. 9.20, I_{max} equals 58 milliamperes and I_{min} is 6 milliamperes. Therefore, the fundamental component of plate current is

$$I_{p_1} = \frac{58 - 6}{2\sqrt{2}} = 18.4 \text{ ma}$$

The alternating plate current flows through an equivalent load resistance of 5,000 ohms. Therefore, the power output is

$$P_0 = I_{p_1}^2 R_L = (0.0184)^2(5{,}000) = 1.69 \text{ watts} \qquad (9.8)$$

This represents power delivered to the transformer; the actual output is less because of transformer losses.

The power supplied to the amplifier by the plate-supply batteries or rectifier is

$$P_{in} = E_b I_b \qquad (9.9)$$

Here I_b is the actual average plate current—more than I_{b_o}. To get an approximate value for the average we can take an average of four equally spaced ordinates because these will always show an average value of zero for both the fundamental and the second-harmonic components. Four convenient ordinates are I_{b_o}, I_{max}, I_{b_o}, and I_{min} because they can all be read from the chart. Therefore,

$$I_b = \frac{I_{max} + I_{min} + 2I_{b_o}}{4} \tag{9.10}$$

Using the values from the load line, we find

$$I_b = \frac{58 + 6 + 2(29)}{4} = 30.5 \text{ ma}$$

The power input is then

$$P_{in} = (0.0305)(250) = 7.63 \text{ watts}$$

This makes the plate efficiency

$$\text{Efficiency} = \frac{(1.69)(100)}{7.63} = 22 \text{ percent}$$

The term *plate* efficiency indicates that only the plate-circuit power has been considered. Including transformer losses and cathode heating power would make the over-all efficiency even worse.

A power output of less than 2 watts at an efficiency of 22 percent does not sound very impressive. More power can be obtained with a triode rated for higher currents and voltages, but the efficiency cannot be improved materially without driving the grid positive to use a larger portion of the available voltage swing. This is seldom done with class A amplifiers.

9.8 The Class A Pentode Power Amplifier

A pentode power amplifier is more efficient than one with triodes because the shape of the characteristic curves permits using a larger portion of the load line. Figure 9.23 illustrates this by showing equivalent triode and pentode amplifiers with the two load lines adjusted for the same current swing and d-c supply voltage. Therefore, the power input to the two amplifiers is approximately the same. However, because of the different shape of the characteristic curves, the alternating plate voltage produced by the pentode exceeds the permissible triode swing. Thus the

pentode circuit has considerably better power output and efficiency than does a triode.

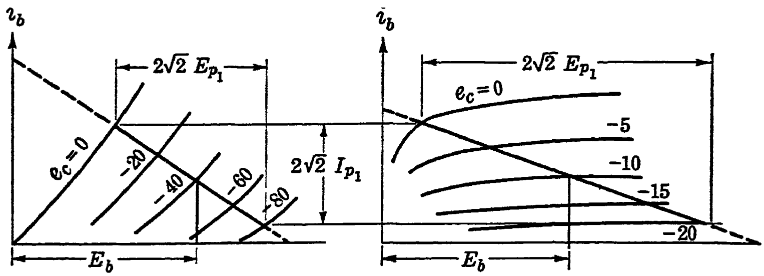


FIG. 9.23. A comparison of equivalent triode and pentode power amplifiers having the same supply voltage and average plate current. The pentode is best because it produces a larger alternating power output for a given d-c input.

As a theoretical upper limit the current and voltage could just swing to zero. Then, neglecting distortion,

$$\sqrt{2}E_p = E_b \qquad E_p = \frac{E_b}{\sqrt{2}}$$

$$\sqrt{2}I_p = I_b \qquad \frac{\quad}{\sqrt{2}}$$

Now computing the efficiency, we obtain

$$\text{Efficiency} = \frac{E_pI_p}{E_bI_b} = \frac{E_bI_b}{2E_bI_b} = 0.5 \text{ or } 50 \text{ percent} \qquad (9.11)$$

The plate efficiency of typical pentode amplifiers usually lies between 35 and 40 percent, with the upper limit for beam-power tubes. The necessity of providing screen-grid current reduces the over-all efficiency to a final value lying between 30 and 35 percent. Again the beam-power tetrode shows its advantage over the conventional pentode because of the low screen current.

The choice of a proper load resistance for a pentode requires a simple analysis, shown by Fig. 9.24. Load line *aa*, for example, represents a low resistance and produces a large current swing but only a small output voltage variation. Owing to the flatness of the pentode curves, a larger load resistance will provide nearly the same alternating current with an even larger voltage swing, and thus a larger output power. Too large a load resistance, how-

ever, brings the upper load-line intersection down to the vertical part of the plate-current curve, as shown by line *cc*. This results in a reduced current swing and serious distortion. Load line *bb*

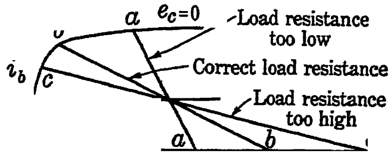


FIG. 9.24. This diagram illustrates the proper choice of load resistance for a pentode power amplifier.

represents a compromise value that passes through the knee of the curve and provides a maximum undistorted voltage swing with a large alternating plate current.

The best load line for a pentode usually results in distortion containing a large amount of third harmonic. The rounding of the characteristic curve knee causes the load-line intercepts to be reduced at both ends of the grid-voltage swing as compared with the middle. This flattens both the top and bottom halves of the output wave form quite differently from a wave containing only second-harmonic distortion, as shown by Fig. 9.22. Third-harmonic content, however, affects both peaks of the wave equally, as suggested by Fig. 9.25. This is generally true of all odd har-

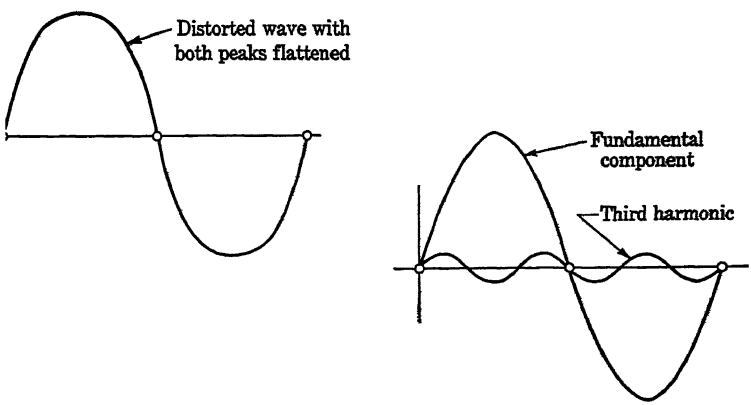


FIG. 9.25. Third-harmonic distortion appears when *both* peaks of the input wave are clipped by the amplifier.

monic frequencies. The distortion produced by pentodes usually exceeds that of triodes, and the high third-harmonic content is considered to be more objectionable than an equal amount of second harmonic in the reproduction of sound.

A power-amplifier pentode requires a smaller signal voltage to drive the grid than does a corresponding triode. Figure 9.23 shows the triode to demand a peak alternating grid voltage of 40 volts to develop full power output, whereas the pentode requires a peak signal of only 10 volts. Since the pentode also exceeds the triode in efficiency, there is little justification for the use of triodes in equipment designed to produce a given power output with the simplest and cheapest circuit design. The power amplifier in a commercial broadcast receiver, for instance, normally employs pentode tubes despite the somewhat worse distortion produced at full output power.

The nicety with which power-amplifier load lines can be constructed and analyzed often leads to a distorted concept of the perfection of the operation of a practical circuit. In many instances the load impedance is only approximately correct because the exact transformer ratio is not available. It may also vary with frequency and be reactive instead of purely resistive. This causes the load line to expand into a distorted ellipse, and the analysis becomes correspondingly complex. The practical result of these deficiencies is to reduce the available power output below the value computed from the ideal resistance-loaded case.

9.9 Push-pull Amplification

A great deal of the distortion produced by curvature of the vacuum-tube characteristic can be eliminated by employing two tubes operating in push-pull, a connection designed to operate one tube in phase opposition to the other. This has the effect of canceling much of the curvature of the two individual characteristics and greatly improves the net linearity. For this reason two tubes in push-pull can provide twice the power obtainable from one tube, but with much less distortion, or more than twice the power with a comparable distortion.

Figure 9.26 shows the basic push-pull circuit. Input transformer T_1 provides two equal voltages of opposite phase with respect to the center tap. This is not the only way of obtaining a phase in-

version; an amplifier designed for an amplification of unity may also provide a phase reversal of 180 degrees. The two push-pull tubes operate from the same bias source with identical grid voltages

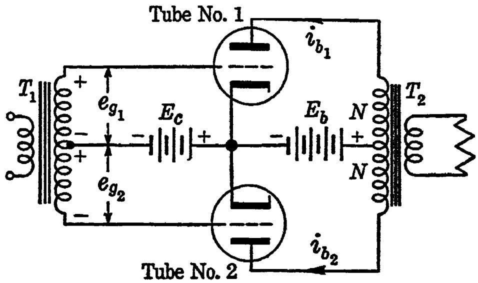


FIG. 9.26. Basic circuit diagram for a push-pull power amplifier.

in phase opposition, and they are supposed to have similar dynamic characteristic curves.

Transformer T_2, with two equal windings of N turns each, combines the output currents in phase opposition. Current i_{b_1} flows up through N turns, and i_{b_2} flows down through an equal number. Since the primary is wound continuously in the same direction with a center tap brought out to divide it into two equal parts, the two currents travel in opposite directions around the iron core and the net magnetization equals the difference between the two. Thus

$$\text{Net magnetization} = Ni_{b_1} - Ni_{b_2} = N(i_{b_1} - i_{b_2})$$

Therefore the output is equivalent to a single current $i_{b_1} - i_{b_2}$ flowing through a single primary of N turns.

Figure 9.27 illustrates the operation of each tube and the combined effect on the output circuit. Figure 9.27*a* shows the operation of tube 1. Although the distortion appears to be extreme, the drawing does not exaggerate the operating conditions in a normal push-pull amplifier. In addition to showing the total plate-current wave, the diagram also shows the fundamental and second-harmonic components of the current. Such a distorted wave, of course, also contains additional harmonic components of appreciable magnitudes.

Figure 9.27*b* illustrates the operation of tube 2, which is identical to that of tube 1 with the exception of the reversed phase of the

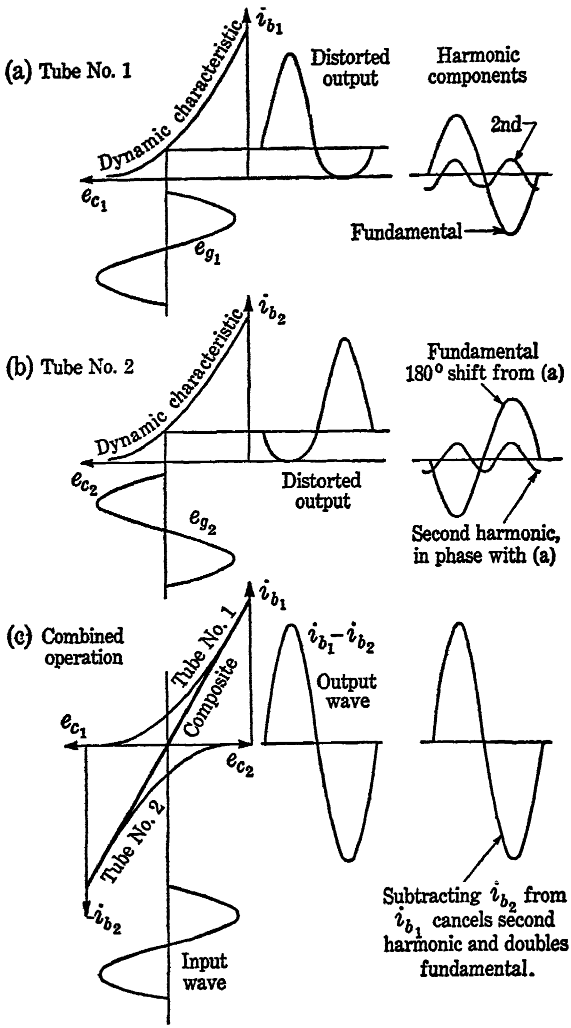


FIG. 9.27. A graphic analysis of the performance of a push-pull power amplifier.

alternating grid voltage. This makes its output wave shape exactly like that of tube 1 except for the 180-degree phase shift. The 180-degree shift also shows in the harmonic components, but since there are two complete second-harmonic waves in one cycle of the fundamental, shifting the second harmonic 180 degrees on

the fundamental scale leaves it apparently in the same position as before.

The net effect, as far as the output transformer is concerned, depends upon the current $i_{b_1} - i_{b_2}$. Thus reversing the wave of *b* and adding it to that of *a* produces the relatively undistorted wave of *c*. Observing the effect of this operation on the harmonic components shows that reversing *b* makes the two fundamental components in phase and the second harmonics out of phase. Therefore the output contains no second-harmonic distortion. In general this process cancels out all even harmonics (second, fourth, etc.) and leaves the odd components (fundamental, third, etc.). The d-c components of the two plate currents also cancel and reduce the magnetization of the transformer core. For this reason push-pull output transformers contain less iron and are less expensive than an equivalent transformer for a single output tube.

In addition to showing the cancellation of harmonics in the output, Fig. 9.27*c* shows a convenient graphic analysis of the situation. The upper curve represents the dynamic characteristic for tube 1 redrawn from part *a*. The curve for tube 2 is also shown, but drawn upside down and back side to. Drawing the curve upside down amounts to reversing the sign on i_{b_2} so that adding ordinates for the two curves will give the current $i_{b_1} - i_{b_2}$. Making the curve also back side to reverses one grid-voltage scale with respect to the other so that the single wave of input voltage will serve for both curves.

Rather than to take two readings from the upper and lower curves and add them each time a point on the output current wave is desired, it is more convenient to add corresponding ordinates to obtain a composite curve. With a properly chosen grid bias, the opposite curvatures of the two tubes can be offset to produce a nearly straight composite dynamic characteristic. A straight line, of course, ensures a linear relationship between the input signal and the net output current and indicates low amplifier distortion.

9.10 General Application of the Push-pull Principle

The push-pull system of reducing nonlinearity applies with equal success to any electrical or mechanical system. For example, an electromechanical measuring device may require greater linearity than can be provided by a simple helical spring. Two such springs

connected in opposition (see Fig. 9.28) will reduce the residual distortion to a very low value.

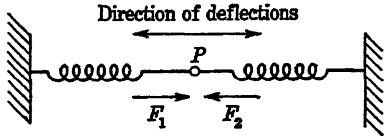


FIG. 9.28. An extension of the push-pull principle to a pair of springs.

To analyze this let us suppose that the smooth curve of Fig. 9.29 represents the nonlinear element to be improved. This curve can be represented by a power series of the following form

$$y = a + bx + cx^2 + dx^3 + ex^4 + \cdots \quad (9.12)$$

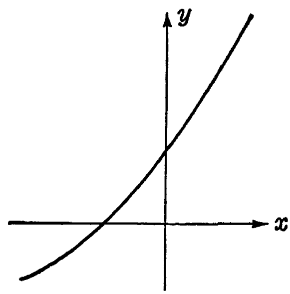


FIG. 9.29. Characteristic curve of a nonlinear device.

With two such elements arranged such that the deflection of one is opposite to that of the other and with their two outputs added subtractively, the linearity can be improved. The arrangement of Fig. 9.28 does this. A deflection causes one spring to compress and the other to stretch, and the net force equals the difference between the two individual spring forces.

Mathematically this amounts to saying that for one element

$$y_1 = a + bx + cx^2 + dx^3 + ex^4 + \cdots$$

For the other element the deflections are opposite in direction. This amounts to replacing $+x$ with $-x$ in Eq. (9.12). Thus

$$y_2 = a - bx + cx^2 - dx^3 + ex^4 - \cdots$$

By taking the difference between y_1 and y_2 to obtain the net output, we get

$$y_1 - y_2 = 2bx + 2dx^3 + \quad (9.13)$$

Therefore the combined output contains only the linear term and all odd powers of x. The improvement obtained depends entirely upon the relative magnitudes of the odd and even terms in the series. Usually, unsymmetrical systems with slight curvature

possess second-order terms large compared with the higher order ones, and the improvement is marked.

Some types of systems possess only odd terms to begin with so that push-pull operation has no advantage. Such an element with a curve of the form

$$y = ax + bx^3 + cx^5 + \cdots$$

must possess the property

$$f(-x) = -f(x)$$

This says that reversing the sign of x changes the sign of the function without changing its magnitude. Figure 9.30 shows a curve of this type. The cantilever beam represents a familiar example of a spring possessing odd symmetry for deflections measured from the normal center position (Fig. 9.31).

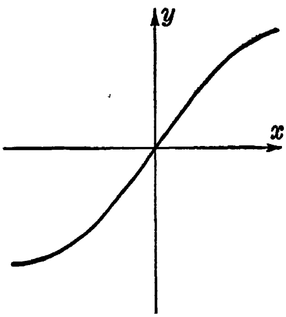


FIG. 9.30. Characteristic curve of a device having odd symmetry.

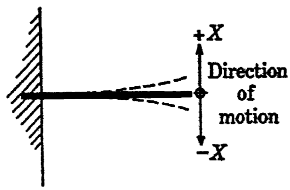


FIG. 9.31. Simple cantilever spring which has a characteristic curve similar to that of Fig. 9.30.

9.11 A Complete Audio-amplifier Circuit

To illustrate the coordination of a number of circuit elements into a complete unit, Fig. 9.32 shows the circuit diagram for a simple 4-watt audio amplifier. This amplifier is designed to amplify the output of a typical phonograph pickup to a level sufficient for operating an ordinary loudspeaker. The circuit provides the functions of volume R_1 and tone controls R_3 and includes a d-c supply rectifier.

The amplifier consists of three separate circuits: (1) a triode amplifier with a gain of about 15 to increase the 1-volt input from

the pickup to a sufficiently large value to drive the 6F6 grid, (2) a pentode power amplifier capable of providing a maximum of 4 watts to the loudspeaker, and (3) a full-wave rectifier to supply the necessary direct voltages to the two amplifier stages. The following descriptions of the function of each circuit element will serve as an aid in understanding the circuit and as a review of the previous chapters on rectification and amplification.

R_1 This serves as the resistor R_g for the 6J5 and as a volume control. The variable tap applies to the grid any desired portion of the input from the phonograph pickup.

R_2 This serves to develop the 4-volt bias required for the 6J5.

R_3 Capacitor C_3 together with R_3 provides the tone control. With R_3 turned to zero, C_3 effectively increases the shunt circuit capacitance enough to make the high-frequency drop occur in the neighborhood of 2,000 cycles. This reduces the needle scratch and produces the commonly desired muted tone. Turning R_3 to maximum resistance increases the impedance so that the shunting effect is negligible and all the frequencies come through.

R_4 This is the plate load resistor R_L for the 6J5.

R_5 The grid resistor of the 6F6 provides a path for any grid current or leakage through C_4.

R_6 Resistor R_6 provides bias for the 6F6 power amplifier.

R_7 This makes up part of an R–C filter in the rectifier and also drops the supply voltage to the correct value for the 6F6 screen and the 6J5 amplifier.

C_1 Capacitor C_1, often omitted, has a negligible reactance to audio frequencies and serves to block off any d-c component in the input signal.

C_2 This serves to by-pass the alternating components of plate current and smooth out the bias provided by R_2.

C_3 C_3 together with R_3 provides the tone control, as discussed in connection with R_3.

C_4 Capacitor C_4 couples the plate of the 6J5 to the grid of the 6F6. It has a low audio-frequency reactance and keeps the direct plate voltage of the first tube from reaching the grid of the second.

C_5 This smooths out the 6F6 bias voltage.

C_6 This is the initial smoothing capacitor for the full-wave rectifier circuit.

C_7 C_7 and L_1 together make up an L–C filter section to provide additional filtering for the voltage supplied to the 6F6 plate.

C_8 C_8 and R_7 provide an additional stage of R–C filter to reduce the hum level before the voltage is applied to the 6F6 screen and 6J5 plate. Since the screen voltage affects a

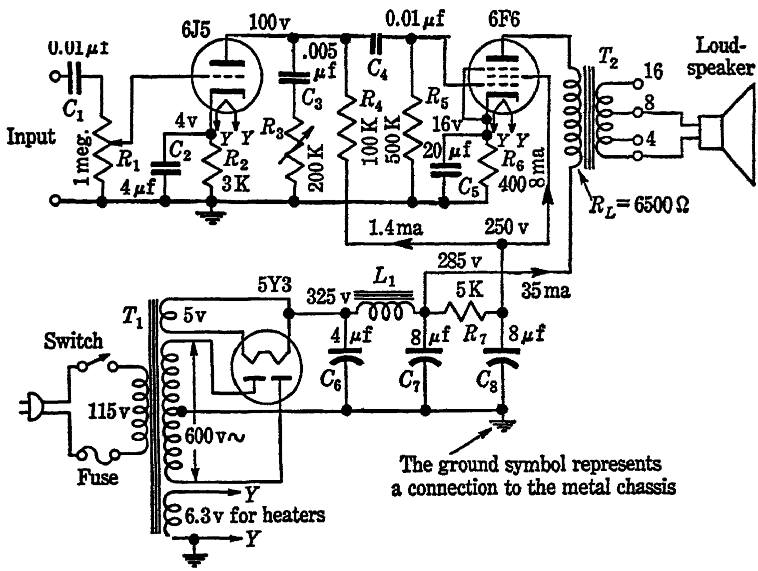


FIG. 9.32. A complete circuit diagram for a simple 4-watt phonograph amplifier.

pentode plate current much more than does the plate potential, the plate supply needs less filtering than does the screen.

T_1 The power transformer provides 600 volts for the full-wave rectifier circuit, 5 volts for heating the rectifier filament, and 6.3 volts (Y–Y) for heating the 6J5 and 6F6 cathodes.

T_2 The output transformer transforms the impedance of the loudspeaker into the correct load for the pentode. The several taps provide for transforming a number of different load impedances into the correct 6,500-ohm load. The diagram shows the correct connection for an 8-ohm loudspeaker.

PROBLEMS

9.1 The circuit of Fig. 9.8 employs a type 6J5 triode with an R_L of 20,000 ohms and an E_{bb} of 300 volts. Determine (*a*) the plate current, and (*b*) the correct size of resistor R_k to provide an 8-volt bias.

9.2 An amplifier for use in certain types of vibration study must have a response that stays flat down to nearly 1 cycle per sec. Compute the approximate size of coupling capacitor required for a grid leak resistance of 1 megohm.

9.3 In a multistage amplifier the response drop at the low-frequency end is cumulative. A drop to 0.9 of the mid-frequency gain in one stage produces a total decrease to 0.81 in two identical stages. Compute the required size of coupling capacitors for a three-stage amplifier in which the over-all gain drops to 0.7 of the mid-frequency value at 5 cycles per sec. The identical stages have grid resistors of 500,000 ohms each.

9.4 An R–C amplifier operates with a high-mu triode (r_p = 100,000 ohms, μ = 100), an R_L of 100,000 ohms, and an R_g of 500,000 ohms for the following tube. Compute the mid-frequency gain.

9.5 An oscilloscope amplifier employs a pentode (g_m = 1,200) in an R–C amplifier with a total shunt capacitance of 40 $\mu\mu$f. What is the maximum possible gain for a cutoff frequency of 200 kc?

9.6 A type 45 triode operates in the circuit of Fig. 9.19 with a plate voltage of 200 volts, a bias of −40 volts, and an effective load resistance of 6,000 ohms. Draw the load line, and determine (*a*) the no-signal plate current, (*b*) the power output, (*c*) the power input and efficiency, and (*d*) the correct transformer turns ratio for connecting the amplifier to an 8-ohm loudspeaker.

9.7 A type 6V6 beam tetrode operates in the circuit of Fig. 9.19 with a plate voltage of 300, a screen voltage of 250, and −12.5 volts of grid bias. Draw the load line for a resistance of 5,000 ohms, and determine (*a*) the approximate power output, (*b*) the approximate power input, and (*c*) the efficiency.

CHAPTER 10

FEEDBACK

In many respects vacuum-tube circuits are unstable devices. The transconductance, amplification factor, and plate resistance of a tube vary with the applied voltage, which makes the amplification of a circuit dependent upon the supply voltage. Tubes themselves are not held to particularly close tolerances, and replacement may markedly change the behavior of a circuit. And finally, all amplifiers produce undesirable distortion.

Many of these deviations from perfection can be reduced by feeding back a small portion of the distorted output voltage to the amplifier input for comparison with the input voltage. This improves the circuit stability, reduces the distortion, and makes the amplifier a precision laboratory device.

10.1 The Feedback Equation

Figure 10.1 shows a line diagram of an amplifier before and after feedback. Before feedback is applied, the input signal E_s is also the grid voltage on the first tube. The amplification A may be

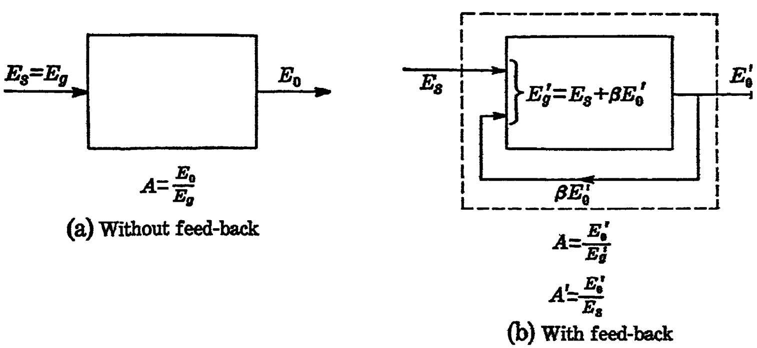
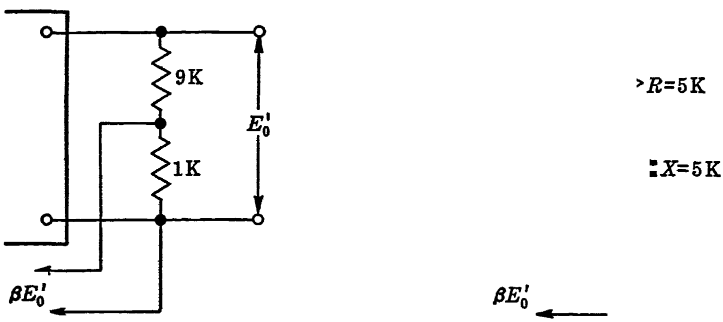


FIG. 10.1. Block diagram of an amplifier before and after the application of a feedback loop.

either a positive or a negative number in the mid-frequency range, depending upon the number of stages of amplification. For a single stage, A is minus because the output and input voltages are in phase opposition, but a two-stage circuit gives a positive value of A. Near the ends of the useful frequency band, phase shifts occur and the amplification includes phase angles other than zero or 180 degrees.

With a feedback loop added to the amplifier, a fraction β of the output voltage returns to the input to add (in the algebraic sense) to the original input signal. It is important to understand that the fraction β not only represents the ratio between the magnitudes of the feedback voltage and the output voltage, but it also includes any phase shift introduced by the feedback circuit. For example, with two resistors determining the feedback, as shown by Fig. 10.2a, β represents the ratio between the voltage across R_2 as



(a) Here $\beta=0.1$ (b) Here $\beta=0.707\angle 45°$

FIG. 10.2. An illustration of the meaning of the fraction β.

compared with E_0', in this case +0.1. With a transformer to reverse the voltage returned to the input, β could be made −0.1. In Fig. 10.2b the fraction fed back also differs in phase from E_0', and β includes a phase angle.

With the feedback loop installed, we are now interested in the new resultant amplification of the over-all system shown inside the dotted line. This we define as A', the new ratio between the output voltage E_0' and the input signal E_s (no longer E_g). Thus

$$A' = \frac{E_0'}{E_s} \tag{10.1}$$

But the ratio between the actual grid voltage and the output voltage is still A

$$E_0' = AE_g = A(E_s + \beta E_0')$$

Solving for E_0', we get

$$E_0' = \frac{AE_s}{1 - A\beta}$$

Then dividing this by E_s, we obtain

$$A' = \frac{E_0'}{E_s} = \frac{A}{1 - A\beta} \tag{10.2}$$

This simple relationship contains a great deal of information; the following discussion points out a few of the salient features involved.

10.2 Negative Feedback

$-A\beta$

Let us first discuss circuits with *negative feedback*, that is, a circuit in which the *product* $A\beta$ is a pure *negative number* over the frequency band of interest. This can never be true at all frequencies because the term A includes phase shifts at one or both ends of the useful frequency band. The negative value of $A\beta$ means that the feedback voltage is out of phase with the input voltage.

Investigating Eq. (10.2) we find that a negative value of $A\beta$ increases the denominator of the expression and makes A' less than A. This reduction of amplification represents the price that must be paid in order to obtain the other advantages of feedback. If now we further make the product $A\beta$ large compared with unity, Eq. (10.2) becomes

$$A' \cong \frac{-1}{\beta} \tag{10.3}$$

Thus the amplification becomes dependent only upon the value of β; variations in A have little effect. If the feedback network consists of two stable resistors, as shown by Fig. 10.2*a*, the fraction β and thus A' will be extremely constant and accurately known.

Negative feedback also decreases the amplitude or nonlinear distortion. A complete analysis of this is difficult, but a qualitative approach is to consider that the distortion results from a variation of the circuit amplification over the operating cycle. When the

tube operates on the steep portion of the curve, A is large; toward the bottom of the curve, A decreases. Since feedback makes A' virtually independent of A, the amplification remains constant over the full cycle and the output voltage accurately follows the input wave.

The amount of correction depends upon the value of $A\beta$. Considering A and A' as variables and β as a constant, Eq. (10.2) in differential form becomes

$$dA' \quad \frac{dA}{(1 - A\beta)^2} \tag{10.4}$$

Repeating Eq. (10.2),

$$A' = \frac{}{1 - A\beta}$$

Dividing Eq. (10.4) by Eq. (10.2),

$$\frac{dA'}{A'} = \frac{dA}{A}\left(\frac{1}{1 - A\beta}\right) \tag{10.5}$$

Although this relation holds exactly for differential changes only, it is approximately correct for small finite changes in A and A'. The equation states that the per unit (multiplied by 100 this would be percent) change in A' equals the per unit change in A reduced by the factor $(1 - A\beta)$. Thus an $A\beta$ of -9 will reduce the effect of variations in A by a factor of 10. If, for instance, normal tube changes and line-voltage variations can change A by 20 percent, negative feedback with an $A\beta$ of -9 will reduce the effect on A' to a 2 percent change.

10.3 Positive Feedback—Oscillation

In a positive feedback circuit the product $A\beta$ is a positive number in the frequency range under investigation. With $A\beta$ less than unity and positive, A' exceeds A; this fact has been used to construct sensitive amplifiers, but the resulting circuit becomes unstable and critical to operate.

An especially interesting condition is obtained when $A\beta$ just equals $+1$. Then

$$A' = \frac{A}{1 - 1} = \infty \tag{10.6}$$

An interpretation of this result is that the amplifier provides an output without any input. Physically it represents a self-excited

amplifier; the fraction of the output fed back to the input is of the proper phase and magnitude just to support the output. Under these conditions the circuit is said to oscillate, and it becomes no longer a useful amplifier. Oscillators in themselves are useful devices, however, for the production of alternating voltages with a flexibility and frequency range unobtainable by rotating machinery.

With $A\beta$ positive and large compared with 1, Eq. (10.2) indicates that A' should again be finite. However, this represents a condition of instability, and oscillations again occur. This can be shown by imagining a small voltage applied to the input. Let us suppose that $A\beta$ equals a positive value greater than unity, say 3. A small voltage E applied to the input produces a voltage AE at the output for a moment, before the signal can return around the feedback loop to the input. A β fraction of this ($A\beta E = 3E$) returns to the input an instant later and adds to the input voltage, making a total input of $4E$. This in turn produces an output $4AE$ and returns through the feedback loop a fraction β, making $4A\beta E$, or $12E$. The input now reaches $13E$, this again returns through the loop to increase the input, and so on. This unstable condition cannot continue indefinitely without running the vacuum tubes to cutoff or up into the positive grid region. As a result, the output increases until the tubes operate with such large signals that the average amplification drops off to make the signal fed back just sufficient to cause a continuous output. This type of oscillation always involves distortion because the nonlinear portion of the characteristic curve limits the output amplitude.

Unfortunately, amplifier oscillation will occur if at *any* frequency the quantity $A\beta$ equals or exceeds a value of $+1$. Thus, although an amplifier may be designed for *negative* feedback over the useful frequency range, sufficient phase shift may occur at some frequency near the ends of the band to make the feedback positive. One way of investigating this is to construct a diagram with $A\beta$ plotted as a vector, showing both its magnitude and phase angle. Figure 10.3 represents such a plot for the amplifier of Fig. 9.10 with two-tenths of the output voltage fed back to the input ($\beta = +0.2$).

At middle frequencies the amplifier gain is -15, which with a β of $+0.2$ provides an $A\beta$ of -3. The diagram shows this as a horizontal vector minus three units long. At the high-frequency

end of the band the amplification drops off and includes a lagging phase angle. At 100 kilocycles, for instance, the gain drops to 13.5 ($A\beta$ drops to 2.7) and the phase shifts about 25 degrees. The diagram also shows the vector plotted for this frequency. In this

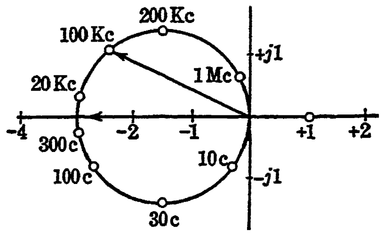


FIG. 10.3. A polar plot of the quantity $A\beta$ for the simple R–C amplifier of Fig. 9.10 with a feedback loop having a β of + 0.2.

way a number of points for different frequencies are obtained, and a curve showing the locus of the $A\beta$ vector is drawn. For this particular amplifier circuit the locus is a circle, and the curve cannot extend into the positive region because the phase shift is limited to plus or minus 90 degrees from the mid-frequency position.

In a two-stage resistance-coupled amplifier the phase shift can approach 180 degrees (90 degrees per stage), and the polar diagram of $A\beta$ becomes something like that shown by Fig. 10.4*a*. Here,

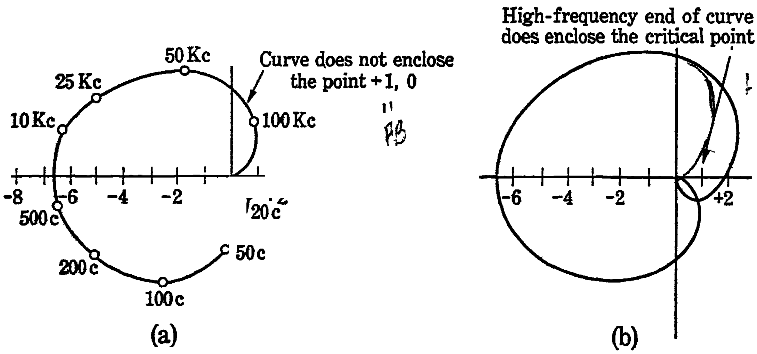


FIG. 10.4. Polar plots showing the locus of $A\beta$ for (*a*) an amplifier that will not oscillate, and (*b*) an amplifier that will oscillate at the higher frequencies.

the curve extends over into the positive region—at 100 kilocycles, for example, the phase shift exceeds 90 degrees although the amplification has only dropped to one-third of the mid-frequency value.

It looks as though the curve might go dangerously near to the +1 point, but it cannot reach it because the amplification drops to zero at the same time that the phase shift approaches 180 degrees. However, in the region where the curve comes close to the critical point, the value of A' may exceed the original amplification A.

This can be seen in the curves of Fig. 10.5 for a two-stage R–C

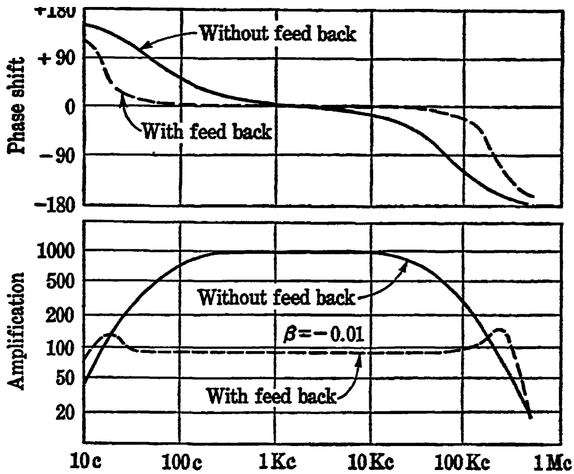


FIG. 10.5. Curves showing the effect of negative feedback in improving the frequency and phase response of a two-stage R–C amplifier.

amplifier. At each end of the frequency band the amplification with feedback actually exceeds the amplification obtained without feedback. The curve clearly shows the advantage of feedback in increasing the frequency band over which the amplification remains constant. The chart also shows the effective reduction of phase shift with feedback.

Figure 10.4*b* shows an even more extreme case of phase shift. This diagram indicates a phase shift approaching 270 degrees at the high-frequency end (this might be caused by a transformer in a two-stage amplifier), and the curve encircles the critical point of +1. Thus although the circuit was designed for negative feedback, at some frequency the feedback is purely positive with $A\beta$ greater than unity and oscillations occur. As a general rule, oscillation always occurs when the polar diagram encircles the critical

One way of avoiding this difficulty is to reduce the value of β

and feed back less of the output voltage. This reduces the size of the polar locus until the intersection with the positive axis falls *inside* the critical point. A lower value of β, however, also diminishes the effectiveness of the feedback in improving the amplifier response in the frequency range of interest.

Probably the most important point to gain from the preceding discussion is that feedback must be applied with care and intelligence. It is not a panacea for the ills of all amplifiers; properly designed circuits offer great improvements, but improper application may cause more harm than good.

10.4 The Implications of Feedback

The feedback principles just discussed need not be limited to amplifiers; they apply equally well to many mechanical and electromechanical control systems which employ a feedback loop to observe the effectiveness of the control. A thermostatically controlled heating system represents a simple example of such a closed control system. A thermostat observes the output of the heating plant and turns off the system when the temperature rises above a predetermined value. The room then cools off, and at a second chosen lower temperature the thermostat again turns on the heat. This type of on-off system must always produce temperature oscillations because it is not continuously temperature sensitive. A better system uses *proportional* control in which the temperature-sensitive control reduces the heating-plant output in proportion to the temperature increase. Properly designed, a system of this type produces a constant temperature for a given set of external conditions, but all too often the system "hunts" and the output temperature oscillates above and below the desired value.

The behavior of a proportional control system can be analyzed with the aid of a polar diagram in exactly the same fashion as that discussed for amplifier feedback. In this particular case, an exact computation of the heat-flow problem would probably be impossible, but the behavior of the feedback loop could be observed experimentally at a number of different frequencies of the heating and cooling cycle. This would require enclosing the temperature-sensitive device and subjecting it to a sinusoidally varying temperature (representing the input to the feedback loop) and measuring the relative amplitude and phase of the room-temperature oscillations (representing the amplifier output). For such a system,

the frequencies of interest would be measured in cycles several minutes long, with the phase shifts normally lagging because of the thermal inertia of the system. The phase difference and the amplitude ratio between the input and output temperature oscillation represent the quantity $A\beta$ to be plotted for each test frequency. At the critical frequency where the phase shift reaches 180 degrees, the value of $A\beta$ must be less than unity; in other words, the amplitude of the room-temperature oscillations must be less than the applied-temperature oscillations. If not, the system will oscillate. The advantage of this type of analysis is that the test data provide the information necessary to correct the system should it be found faulty. Inspection of the curve may show how the phase shift can be corrected or the feedback reduced to stabilize the operation.

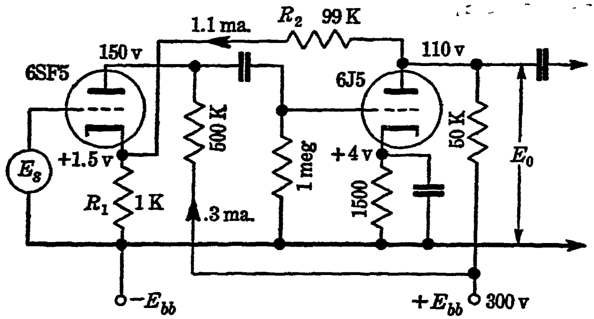
If the system includes an electrical link, it may be simpler to break the loop at this point so that the input and output can be observed electrically. For example, one electromechanical system for measuring the force on wind-tunnel models employs a mechanical linkage operating a beam whose position is observed electrically with the aid of a pair of phototubes. The amplified photocell output in turn operates an electromagnet to produce a restoring force that maintains the beam in the balance position. Since the restoring force is proportional to the electromagnet current, a simple ammeter reading provides a direct indication of the original force on the model. As originally constructed the system oscillated severely, but the installation of damping dashpots corrected the difficulty. From the standpoint of feedback theory this was equivalent to adding resistance to the system and reducing the phase shift below 180 degrees until the amplification around the loop became less than one.

To observe this, the circuit could be opened between the photocells and the amplifier and small alternating voltages of different frequencies applied to the amplifier input. The electrical photocell output compared with this input provides the data for the polar-diagram analysis. In this particular case the frequencies of interest lie in the vicinity of 1 cycle per second.

10.5 A Practical Feedback Circuit

There are many different types of feedback circuits, but Fig. 10.6 shows one of the simplest and most practical arrangements for

applying a portion of the output voltage to the input circuit. This circuit takes advantage of the fact that applying the feedback to the cathode instead of to the grid of the input tube amounts to



$A = +1000$ $\beta = -0.01$ $A' = 91$

FIG. 10.6. A practical feedback amplifier showing a portion of the output voltage applied to the cathode of the input tube. This is equivalent to applying an equal voltage of opposite polarity to the grid.

inverting the phase of β. It also makes the circuit design simple; applying voltage feedback directly to the grid usually involves circuit complications.

Resistors R_1 and R_2 effectively form a voltage divider across the output of the second tube, and a fraction $R_1/(R_1 + R_2)$ of the output voltage returns to the input cathode. This voltage is in phase with the output voltage. For negative feedback the quantity $A\beta$ must be negative, and since A is positive for a two-stage amplifier, factor β must carry the negative sign. Applying the feedback voltage to the cathode accomplishes this phase reversal because raising the cathode voltage is equivalent to lowering the grid potential.

In addition to providing the feedback, resistor R_1 also produces the grid-bias voltage for the input tube. In Fig. 10.6 the circuit values are carefully worked out so that the plate current of the first tube plus the direct current flowing through R_2 combine in R_1 to provide the correct bias voltage. In many instances the circuit constants cannot be juggled to permit this, and a blocking condenser must be inserted to remove the d-c component of current through R_2. With this arrangement the designer can first choose R_1 to provide correct bias, then compute R_2 for the correct value of feedback. The blocking condenser, of course, increases the

phase shift in the feedback loop and decreases the useful frequency range.

10.6 The Cathode Follower

The unusual amplifier of Fig. 10.7 represents a circuit with so much feedback that the net amplification becomes less than unity.

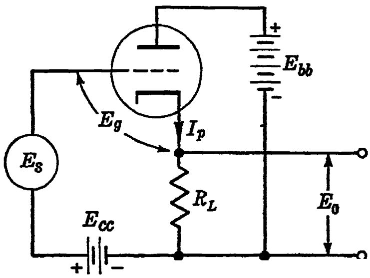


FIG. 10.7. The basic cathode-follower circuit.

(Perhaps the term "damplifier" would be appropriate.) Although this may seem as anomalous as a bird without wings, yet the cathode follower, as it is called, possesses definite characteristics useful in instrumentation.

Qualitatively, the circuit operation is easy to understand. Increasing the input voltage E_s in the positive direction increases the plate current, which in turn increases the drop in R_L and raises the cathode potential. Lowering the input voltage drops the plate current and lowers the cathode voltage. Thus the cathode follows the input voltage up and down, and the output remains in phase with the input. Hence the term cathode follower. The positive bias voltage E_{cc} offsets some of the voltage drop in R_L to provide the correct net operating grid bias because the drop in R_L usually exceeds the required bias.

A convenient analysis of the circuit is to consider this as an amplifier in which the load resistance R_L has been moved from the plate around to the cathode circuit. In this position the same plate current passes through R_L as in a conventional amplifier, and the equation for the amplification $A(E_0/E_g)$ before feedback remains the same as for a conventional triode circuit except for a change in sign.

$$A = \frac{+\mu R_L}{r_p + R_L} \tag{10.7}$$

The change in sign occurs because here E_0 equals the voltage drop in R_L directly without subtraction from the supply voltage E_{bb}.

This arrangement makes β unity because the full output voltage E_0 appears in the cathode circuit. Factor β is also negative because inserting $+E_0$ in the cathode circuit amounts to inserting $-E_0$ in the grid circuit, as discussed in the previous section.

By placing this value for β in the feedback equation (10.2), we obtain

$$A' = \frac{A}{1 - (-1)A} = \frac{A}{A + 1} \tag{10.8}$$

Since A normally exceeds 10, Eq. (10.8) shows that A' closely approaches but always remains less than unity. Also A' is very stable; changing A from 30 to 40 changes A' from 30/31 to 40/41, a difference of only 0.8 percent.

The prime advantage of the cathode follower is its extreme stability, due to the large amount of feedback, and its ability to transfer the input voltage E_s, from which it draws no power, over to the load resistance R_L with only a slight loss in magnitude and with negligible distortion. This permits the construction of a vacuum-tube voltmeter, for example, which draws no power from the voltage being measured, but which accurately transfers the input voltage changes over to a measuring circuit consisting of R_L in series with a meter.

We shall now develop an equivalent circuit for the cathode follower somewhat similar to the equivalent circuit for a triode. In so doing we shall consider R_L as a load resistor separate from the cathode follower itself and obtain an expression for E_0 in terms of I_p. First we replace A in Eq. (10.8) by Eq. (10.7)

$$A' = \frac{\mu R_L}{r_p + (\mu + 1)R_L} \tag{10.9}$$

But A' represents the ratio between input and output voltages; $E_0 = A'E_s$. With A' replaced by expression (10.9), we get

$$E_0 = \frac{\mu R_L E_s}{r_p + (\mu + 1)R_L}$$

Rearranged to collect terms involving R_L, the equation becomes

$$R_L[(\mu + 1)E_0 - \mu E_s] = -r_p E_0 \tag{10.10}$$

But R_L equals E_0/I_p. Therefore,

$$\frac{E_0}{I_p}[(\mu + 1)E_0 - \mu E_s] = -r_p E_0 \tag{10.11}$$

After solving Eq. (10.11) for E_0, we obtain

$$E_0 = \left(\frac{\mu}{\mu + 1}\right)E_s - \left(\frac{r_p}{\mu + 1}\right)I_p \tag{10.12}$$

Since μ is a dimensionless ratio, the first term following the equality represents a voltage and the second term an IR drop. Figure 10.8

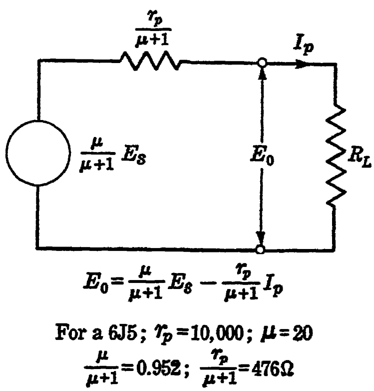


FIG. 10.8. The equivalent circuit of a cathode-follower amplifier.

displays this relationship in a more striking manner by means of a simple series circuit showing a voltage $\mu E_s/(\mu + 1)$ in series with a resistance $r_p/(\mu + 1)$. Inspection will verify that this circuit follows Eq. (10.12) and is thus equivalent to the cathode follower of Fig. 10.7.

This equivalent circuit shows that a cathode follower acts to translate the input voltage, from which it draws no power, over to an equivalent generator with a low internal resistance from which power can be drawn. The circuit achieves this with little loss in voltage and with considerable stability because of the negative feedback employed.

The equivalent circuit also simplifies the design of equipment employing a cathode follower. The vacuum-tube voltmeter discussed in Chap. 16 illustrates an application of this.

PROBLEMS

10.1 An amplifier with an output voltage of 30 volts feeds back a voltage of 2 volts lagging the output by 30 deg. Compute β.

10.2 The mid-frequency gain of a two-stage amplifier is +1,000. (*a*) Compute A' for the amplifier with a feedback loop providing a β of −0.01. (*b*) Recompute A' under the assumption that changing tubes changes A to a new

value of 800. (c) Compare the per unit change in A' to the per unit change in A.

10.3 An amplifier with a gain of $800\underline{/30°}$ is provided with a feedback loop in which β is -0.01. Compute A', and compare the new phase shift with the original value.

10.4 The table below shows the amplification data for a two-stage amplifier including the phase shifts due to the cathode bias and screen-dropping circuits. (a) Plot a polar diagram for the factor $A\beta$ if β has a constant value of -0.02. (b) What is the maximum allowable magnitude of β if the feedback circuit involves no phase shifts?

Frequency, cps	A	Frequency, kc per sec	A
10	$21\underline{/260°}$	10	$2{,}800\underline{/-25°}$
20	$120\underline{/200°}$	20	$2{,}400\underline{/-45°}$
50	$850\underline{/145°}$	50	$1{,}500\underline{/-85°}$
100	$1{,}900\underline{/100°}$	100	$650\underline{/-115°}$
200	$2{,}700\underline{/55°}$	200	$190\underline{/-140°}$
500	$3{,}000\underline{/16°}$	500	$34\underline{/-165°}$
5,000	$3{,}000\underline{/-10°}$		

CHAPTER 11

RESONANCE AND TUNED AMPLIFIERS

Resonant circuits and tuned amplifiers play an important role in the field of electronics. They provide the selectivity required to select a single signal from a group of signals of different frequencies, whether they are radio broadcasts or telephone messages on a long-distance line. Frequency selective circuits may greatly increase the sensitivity and reliability of a control system by distinguishing between the desired frequency and unwanted interference and background noise. Radio-frequency oscillators used to provide high-frequency electrical energy for high-speed heat-treatment, wood gluing, and plastic molding are essentially tuned amplifiers with positive feedback. From an engineering standpoint, tuned power amplifiers are particularly important because they can be designed for efficiencies above 75 percent—high enough to merit the respect of a mechanical engineer.

11.1 Series Resonance

Figure 11.1 shows a simple series resonant circuit. The impedance of this circuit is

$$Z = r + j(X_L - X_c) \qquad (11.1)$$

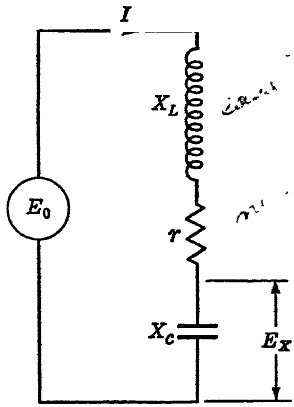


FIG. 11.1. Simple series resonant circuit.

Here r is the total circuit resistance, usually only that of the inductance alone to keep the resistance low. Terms X_L and X_c represent the magnitudes of the inductive and capacitive reactances in the circuit. Since one increases with frequency ($X_L = 2\pi fL$) and the other decreases ($X_c = 1/2\pi fC$), there is always some particular frequency at which the two become equal and cancel. This defines the resonant frequency. At this point the impedance reaches a

minimum and the current a maximum. At resonance

$$Z_r = r + j0 = r \tag{11.2}$$

$$I_r = \frac{E_0}{Z_r} = \frac{E_0}{r} \tag{11.3}$$

The subscript r refers to the resonant condition. We can find the resonant frequency by equating X_L and X_c, the condition of resonance.

$$2\pi f_r L = \frac{1}{2\pi f_r C}$$

$$f_r = \frac{1}{2\pi\sqrt{LC}} \tag{11.4}$$

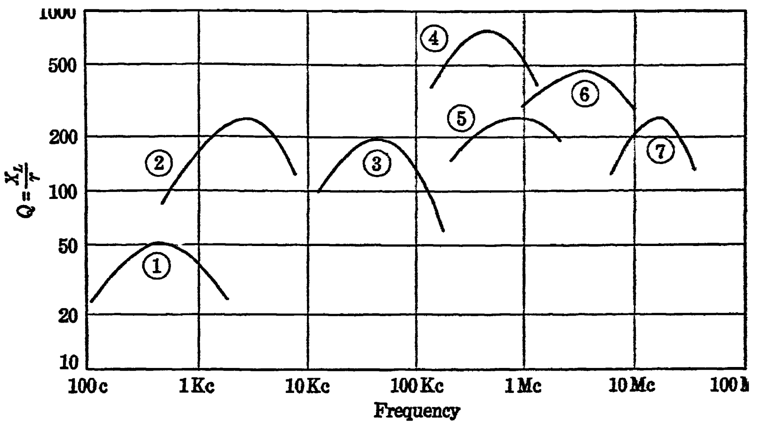
Factor Q. An important factor controlling the behavior of a resonant circuit is the ratio between the reactance and the resistance. This factor, called Q, is defined by the expression

$$Q = \frac{X_L}{r} = \frac{2\pi f L}{r} \tag{11.5}$$

The reason for comparing the reactance of the inductor to the circuit resistance is that the coil provides the majority of this resistance because practical capacitors are nearly lossless. With particular care to keep the resistance low, the Q of a coil may be several hundred at moderate radio frequencies.

The curves of Fig. 11.2 show the factor Q for a number of typical coils over a wide frequency range. Each curve shows a broad maximum and then drops off instead of increasing indefinitely with the frequency, as might be expected from Eq. (11.5). This happens because the effective coil resistance also increases with the frequency, slowly at first, but finally even more rapidly than the reactance. In iron-cored coils this results from the rapid increase in hysteresis and eddy-current losses in the core as the frequency is raised. In air-core coils the high-frequency skin effect which crowds the current to the conductor surface reduces the effective cross-sectional area and increases the resistance.

At low audio frequencies the low inductive reactance of a given coil makes it difficult to obtain a Q of above 50. Laminated iron-cored coils usually provide better performance for these frequencies than a coil of a similar physical size without iron. In the high audio- and low radio-frequency region the use of powdered-iron cores reduces the size of coil needed for a given inductance and permits obtaining rather high values of Q. These cores consist



(1) L = 1 h, laminated iron core
(2) and (3) L = 20 mh, permalloy dustcore
(4) L = 500 μh, air core, Litz (Stranded)
(5) L = 300 μh, air core, No. 16 Cu wire
(6) L = 10 μh, air core, No. 10 Cu wire
(7) L = 2 μh, air core, No. 16 Cu wire

FIG. 11.2. Curves showing typical values of factor Q at different frequencies for a number of different inductors.

of high-permeability iron alloy reduced to a fine powder, mixed with plastic, and molded to shape. This gives the equivalent of extremely fine laminations and reduces the eddy-current loss.

For frequencies above about 1 megacycle, simple single-layer coils wound on cylindrical forms perform as well as more complex windings, although many different forms have been tried. Most communications handbooks give tables for computing the coil size and winding required to produce a given inductance.

At intermediate frequencies between 100 kilocycles and 1 megacycle, either air-core or iron-dust-cored coils produce equally good results; the iron cores have the advantage of reducing the physical size of the inductor. In fact, some radio receivers are tuned by moving powdered-iron slugs into and out of the windings.

Resonant Voltage Rise. It is especially interesting to observe the value of the voltage E_x across one of the reactances at resonance.

$$E_x = X_c I_r$$

But from Eq. (11.3) I_r is E_0/r, and at resonance X_L equals X_c. With these substitutions, we obtain

$$\boxed{E_x = \frac{E_0 X_L}{r} = QE_0} \qquad (11.6)$$

This shows that the voltage across the condenser may be several hundred times the applied voltage at the resonant frequency.

However, this differs from the voltage step up of a transformer because any attempt to connect a load in parallel with X_c changes the circuit, throws it out of resonance, and reduces E_x to a small value. This resonant rise can be used in a vacuum-tube amplifier, however, because a vacuum tube requires no input power and does not disturb the resonant circuit.

Off resonance the voltage drops rapidly and becomes even less than the applied voltage for frequencies sufficiently removed from resonance.

Selectivity. At frequencies either side of resonance the capacitive and inductive reactances do not cancel [see Eq. (11.1)], the circuit impedance increases sharply, and the current drops rapidly, as shown by Fig. 11.3. Thus the circuit responds best to one particu-

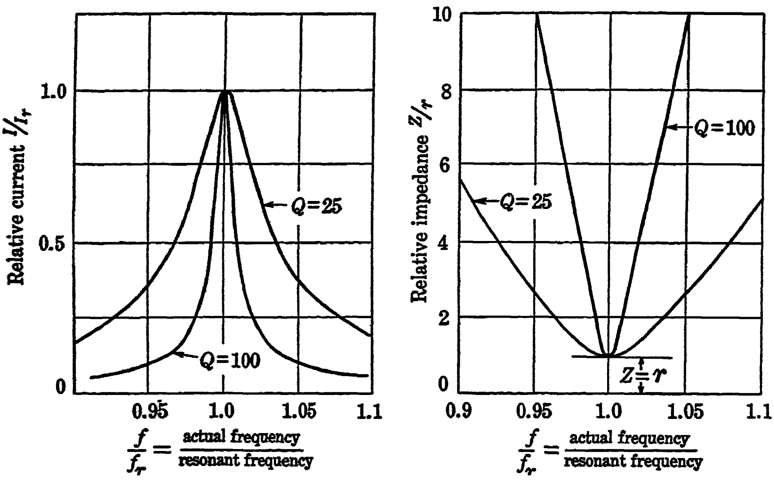


FIG. 11.3. Curves showing the selectivity characteristics of a series resonant circuit for two values of Q.

lar frequency (or narrow band of frequencies) and poorly to frequencies removed from the resonant value. For this reason the circuit is said to be frequency selective.

Selectivity represents the relative ability of a circuit to select one frequency out of a number of frequencies in the applied voltage. Figure 11.3 shows this property graphically for two different values of Q. With a Q of 100 the circuit is quite sharply selective, and it discriminates against a frequency 5 percent away from resonance by a factor of 10. A Q of only 25 makes the curve

blunter and provides a selectivity about one-fourth as good—for a given ordinate the distance between two points on the curve is four times larger than for a Q of 100.

11.2 Parallel Resonance

Amplifiers usually employ parallel resonant circuits similar to that shown by Fig. 11.4. This circuit has selectivity properties similar to a series resonant circuit, but in many ways it behaves in reciprocal fashion; resonance produces minimum current and maximum impedance.

The diagram of Fig. 11.4 actually shows a series-parallel circuit

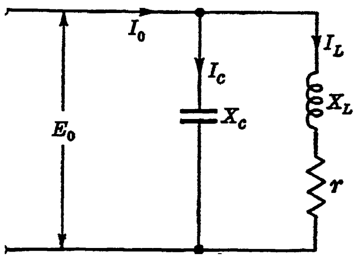


FIG. 11.4. A parallel resonant circuit.

because the resistance appears in series with the inductance (assuming no resistance other than that of the coil itself). For this reason the analysis of its operation becomes more complex, and we shall content ourselves with an approximation. Figure 11.5*a* shows a vector diagram of the parallel circuit at resonance. Current I_c leads the applied voltage by 90 degrees and current I_L

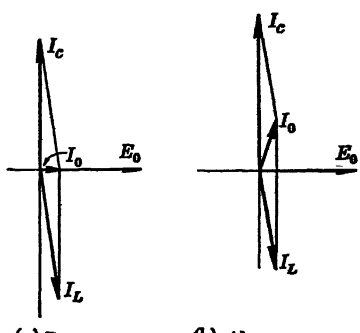


FIG. 11.5. Vector diagram for the parallel circuit of Fig. 11.4.

lags by an angle approaching 90 degrees more closely than indicated by the diagram. At a particular frequency where I_c and the vertical component of I_L (practically equal to I_L itself) become equal, the current I_0 reaches a very small value in phase with the applied voltage and the circuit appears as a high resistance R to the voltage source. This contrasts with the series case which appeared as a very small resistance at resonance.

At frequencies different from resonance, I_L and I_c no longer balance and I_0 increases, as shown by Fig. 11.5*b*. The impedance therefore drops off rapidly on either side of the resonance point, as shown by Fig. 11.6 for two values of Q. As with the series

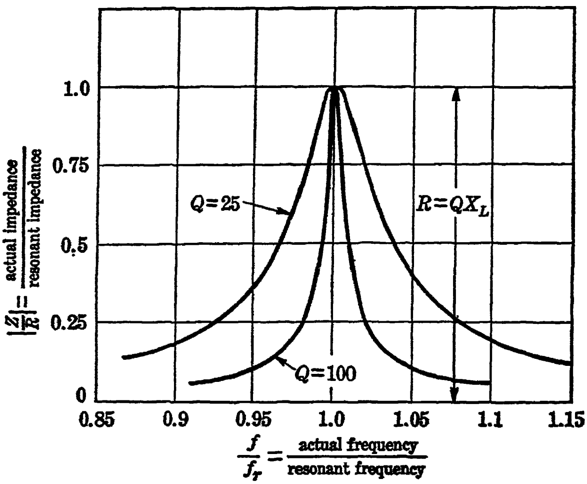


FIG. 11.6. Impedance curves illustrating the selectivity characteristics of a parallel resonant circuit.

circuit the higher Q produces the sharper resonance curve; the width of the curve at any point near resonance is nearly inversely proportional to Q.

To compute the impedance of the parallel circuit at resonance, we use the usual inverse rule for parallel circuits

$$\frac{1}{Z}=\frac{1}{r+jX_L}+\frac{1}{-jX_c}$$

$$\frac{1}{Z}=\frac{r-jX_L}{r^2+X_L^2}+\frac{j}{X_c}$$

$$\frac{1}{Z}=\frac{r}{r^2+X_L^2}+j\left(\frac{1}{X_c}-\frac{X_L}{r^2+X_L^2}\right) \tag{11.7}$$

The second equation came from multiplying the numerator and denominator of the first term by the conjugate $(r - jX_L)$.

At resonance the circuit impedance becomes purely resistive. For this to be so the second term of Eq. (11.7) must vanish to leave no reactive component. Therefore,

$$\frac{1}{X_c} = \frac{X_L}{r^2 + X_L^2} \tag{11.8}$$

For a Q greater than 10 the quantity X_L^2 exceeds r^2 by a factor of 100, and with little error we can neglect the r^2 terms in Eq. (11.8). This gives as the condition of resonance

$$\frac{1}{X_c} \cong \frac{1}{X_L}$$

$$X_c \cong X_L$$

This corresponds to the definition of resonance for the series circuit and therefore the resonant frequency is given closely by

$$f_r \cong \frac{1}{2\pi\sqrt{LC}} \tag{11.4}$$

We can obtain an approximate value for the impedance at resonance by neglecting r^2. From Eq. (11.7) with the j component equal to zero at resonance, we obtain

$$\frac{1}{Z_r} \cong \frac{r}{X_L^2}$$

$$Z_r \equiv R \cong X_L^2/r = QX_L \tag{11.9}$$

This shows that the parallel resonant impedance greatly exceeds the circuit reactance, although it drops off rapidly on either side of the resonant frequency.

11.3 Function of the Resonant Circuit in a Tuned Amplifier

Figure 11.7 shows the circuit diagram for a simple tuned amplifier using a pentode. A pentode reduces the feedback by practically

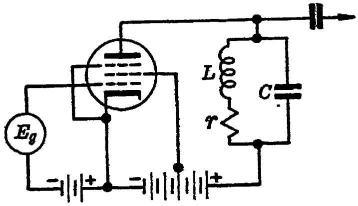


FIG. 11.7. Basic circuit diagram for a tuned amplifier.

eliminating the grid-plate capacitance, and it also provides more amplification than does a triode. The only difference between the circuit of Fig. 11.7 and a normal pentode audio amplifier is the substitution of a parallel resonant circuit for the load resistance. In fact at resonance the circuit operates exactly as though it had a resistance load equal to QX_L because at this frequency the parallel circuit presents a pure resistance to the tube.

Selectivity. From Eq. (4.19) giving the amplification of a pentode circuit

$$A = -g_m Z_L = -g_m Q X_L \qquad (4.19)$$

Therefore the amplification curve shown by Fig. 11.8 exactly follows the impedance curve of Fig. 11.6. With a reasonably high Q

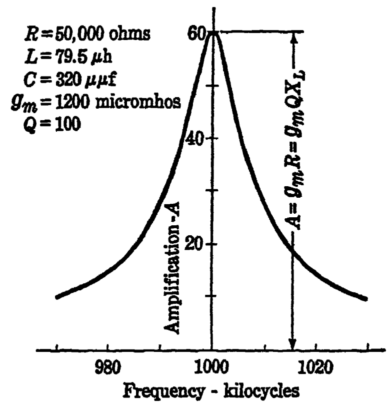


FIG. 11.8. Frequency-response curve for the tuned amplifier of Fig. 11.7.

the circuit amplifies a narrow band of frequencies well, but other signals differing appreciably from the resonant frequency receive little amplification. This provides the selective amplification used in radio receivers to select one particular station from a signal containing a large number of programs transmitted at different frequencies. A single tuned circuit cannot ordinarily accomplish this; the process is repeated three or more times to obtain sufficient selectivity.

Increased Load Impedance. At high frequencies a tuned circuit is necessary to obtain a reasonable amount of amplification despite the need for selectivity. The discussion of the R–C amplifier points out that unavoidable shunt capacitance makes it difficult to obtain a reasonably large load impedance at the higher fre-

quencies. For example, the reactance of a 30-micromicrofarad capacitance at 10 megacycles is but 530 ohms, and this in shunt with a load resistor produces an even lower value of load impedance. With a typical transconductance of 1,500 micromhos the amplification computed from Eq. (4.19) turns out to be less than unity—hardly a satisfactory state of affairs. By making this stray capacitance part of the tuning capacitance of a parallel resonant circuit, however, we can obtain a large load impedance and correspondingly high gain. With a total capacitance of 80 micromicrofarads (30 micromicrofarads plus a variable condenser of 50 micromicrofarads) in parallel with a coil having a Q of only 100, the resonant load impedance becomes QX_L, or 20,000 ohms when tuned to resonance at 10 megacycles. A load impedance of this magnitude will provide a gain of about 30 with typical pentodes used in class A tuned amplifiers.

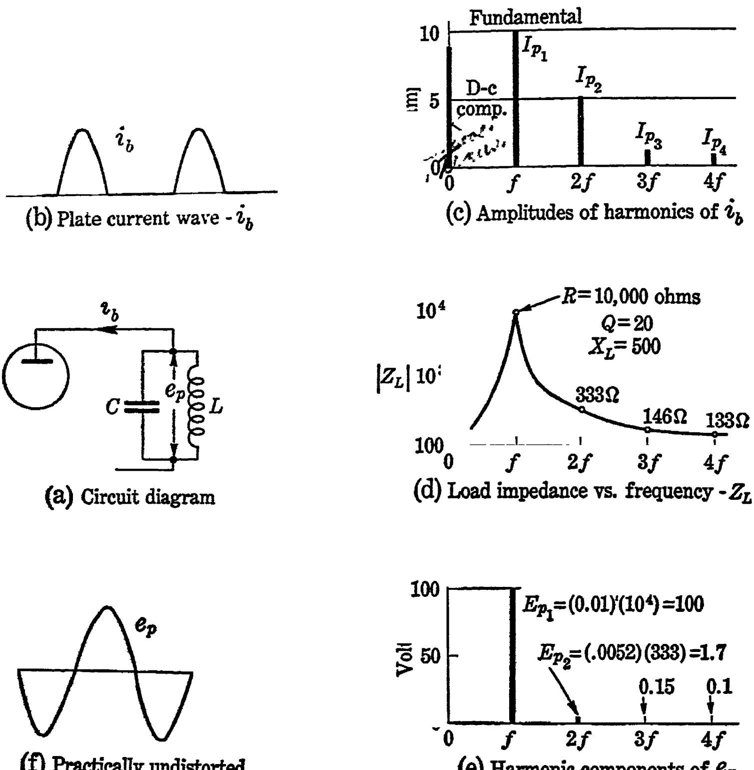
Restoration of Wave Form. In audio power amplifiers the most effective design strikes a compromise between maximum output, good efficiency, and minimum distortion. A tuned power amplifier, however, can amplify a single frequency only (or very narrow band), and the selective properties of the tuned circuit practically eliminate any amplitude distortion produced by the tube itself. This permits operating the circuit with little regard for distortion to obtain plate-circuit efficiencies as high as 80 percent.

Figure 11.9 shows a quantitative analysis of this wave-form restoration. We shall assume the vacuum tube to be biased beyond cutoff (class C operation) with a large sinusoidal voltage applied to the grid. Plate current flows less than half the time and the plate-current wave form looks like the tips of sine waves shown by Fig. 11.9*b*. A harmonic analysis of this terribly distorted wave (Fig. 11.9*c*) shows a second harmonic more than half the size of the fundamental, with appreciable third and fourth harmonics.

In Fig. 11.9*d* a curve of load impedance against frequency shows the characteristics of the load impedance through which the plate-current pulses pass. This tuned circuit presents a large impedance to the fundamental component of plate current (I_{p_1}), but the impedance to the harmonics is very much smaller, even with a circuit Q as low as 20.

Figure 11.9*e* shows the different harmonic components of the resulting plate voltage. Each is computed from the product of the

corresponding plate-current component by the load impedance at that particular frequency. Thus, although the plate current contains about 50 percent of second harmonic, the low load impedance at this frequency produces a second-harmonic voltage drop less



(b) Plate current wave - i_b

(c) Amplitudes of harmonics of i_b

(a) Circuit diagram

(d) Load impedance vs. frequency - Z_L

(f) Practically undistorted output voltage - e_p

(e) Harmonic components of e_p computed from (c) and (d)

FIG. 11.9. This series of diagrams illustrates how the resonant load circuit in a tuned amplifier restores the sinusoidal output wave form despite distortion of the plate-current wave.

than 2 percent of the fundamental output voltage. The higher harmonics drop completely out of the picture, and the output voltage wave of Fig. 11.9*f* looks practically distortionless.

This resonant circuit has a mechanical model in the clock pendulum energized by force pulses, yet which oscillates back and forth with nearly sinusoidal motion. The pendulum provides the

energy storage required to smooth out the input pulses into simple harmonic motion. This storage takes place in two forms, kinetic energy which is maximum when the velocity is greatest, and potential energy which reaches a maximum at the ends of each swing. Similarly, in the tuned circuit the storage occurs in the magnetic and electric fields of L and C. At the moment of peak voltage and zero current the energy is all stored in the capacitance, while at the current maximum the magnetic field receives the energy.

11.4 Tuned Power Amplifiers

The ability of a tuned circuit to restore the sinusoidal wave shape simplifies the problem of obtaining high-efficiency operation because plate-current distortion no longer limits the design.

Plate Efficiency. The plate efficiency of both the triode and pentode amplifiers discussed in Chap. 9 was poor. Figure 11.10

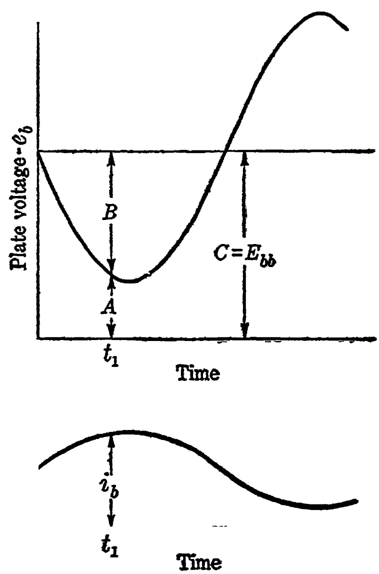


FIG. 11.10. An aid in the analysis of amplifier-plate efficiency.

aids our analysis of this fact. At an instant t_1 the ordinate C represents the plate-supply voltage because the d-c drop in the tuned circuit is negligible. At this same instant B represents the voltage drop across the load impedance, while A is the voltage

drop across the tube itself. The momentary power relations are

$$\text{Power input} = C(i_b)$$
$$\text{Power loss in tube} = A(i_b)$$

The power output to the load equals the input minus the loss.

$$\text{Power output} = C(i_b) - A(i_b) = B(i_b)$$

The *instantaneous* efficiency is therefore

$$\text{Efficiency} = \frac{\text{output}}{\text{input}} = \frac{B(i_b)}{C(i_b)} = \frac{B}{C} \tag{11.10}$$

At the moment shown by Fig. 11.10 the efficiency is about 60 percent; at the plate current peak it is even higher, but away from this optimum point the efficiency drops off rapidly. The *average* efficiency, of course, must be obtained by properly weighting and averaging the instantaneous values, but in no case can the average efficiency exceed the peak instantaneous efficiency. Fortunately, however, it may approach the maximum because the plate current is largest and the maximum amount of power flows at the efficiency peak. This immediately suggests two ways of improving the amplifier performance: (1) operate with as large a plate-voltage swing as possible to make B approach C at the bottom of the swing; (2) allow plate current to flow only for a short time during the period of high efficiency. These considerations dictate the choice of a class C amplifier for efficient power amplification of a single frequency.

Class C Wave Forms. Although the circuit diagram of Fig. 11.11 for a class C amplifier looks just like the diagram for a class A circuit, the important difference is in the relative sizes of the voltages applied. For class C operation, bias voltage E_c exceeds cutoff and plate current flows for less than half a cycle of the applied alternating grid voltage. Thus power flows only during the period of high efficiency, which keeps the average efficiency up to 75 percent or more. To obtain this good efficiency the peak value of the alternating grid voltage must exceed the grid bias to make the net grid voltage positive during most of the conducting period. This aids the plate in attracting the electron stream and allows a large plate current to flow with a low plate voltage.

Driving the grid positive produces grid current flow, as shown by the wave forms of Fig. 11.11. This necessitates a power ampli-

fier to provide the grid driving power, but this is usually a small price to pay for the good efficiency obtained in the class C stage. The effect of grid current in distorting the input voltage wave is

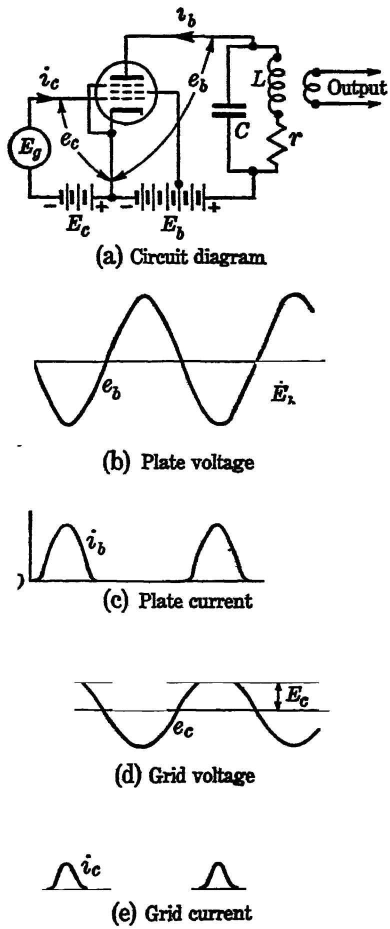


FIG. 11.11. Circuit diagram and wave forms for a class C tuned amplifier.

unimportant because the output tuned circuit removes the harmonics and saves only the fundamental component.

Power Output. The inductively coupled output coil of Fig. 11.11 suggests one method of removing the power output of the amplifier. Since this circuit contains no iron, only a small fraction of the flux produced by L links the output coil. Consequently, the

approximations involved in developing the transformer-impedance-matching concept of Chap. 9 no longer hold. A more convenient concept is to imagine that the power actually transferred to the secondary is absorbed in an equivalent resistance in series with L. This reflected resistance makes the effective Q of the circuit less than that of the inductor itself, and resistor r represents the actual coil resistance plus the reflected resistance.

To obtain high efficiency in transferring energy to the secondary, the primary Q must be as high as possible to minimize the true coil resistance. The coupling is then designed to make the effective circuit Q (X_L divided by the sum of the coil and the reflected resistance) as low as permissible (about 12). This makes r relatively large, and the coil resistance becomes small by comparison.

Class C Limitations. Despite the attractive efficiency of a class C circuit its application is distinctly limited to cases involving the amplification of a *single* frequency of *constant* amplitude. The tuned circuit imposes the frequency restriction, and the extremely large bias imposes the second. With a bias beyond cutoff, alternating input signals below a limiting value produce no output. Figure 11.12 shows this condition. Above the limiting value the

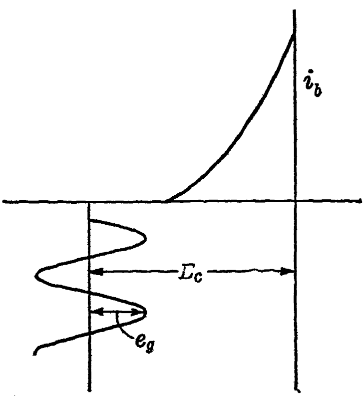


FIG. 11.12. Class C amplifier with insufficient input signal.

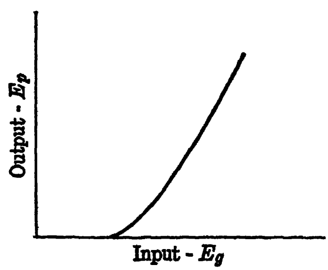


FIG. 11.13. A curve showing the nonlinear relation between the effective input and output voltages of a class C amplifier.

relationship between the input and output amplitudes is not linear; the curve usually takes the shape of Fig. 11.13. For this

reason it cannot properly amplify the modulated voltages discussed in Chap. 13.

These restrictions make the class C amplifier appear to be of limited usefulness. However, all radio transmitters employ them, and most oscillators for providing radio-frequency power are essentially self-excited class C amplifiers.

Grid-leak Bias. The term grid-leak bias refers to a system of obtaining the grid bias by passing the grid current through a resistor, as shown by Fig. 11.14*a*. A convenient analysis is to

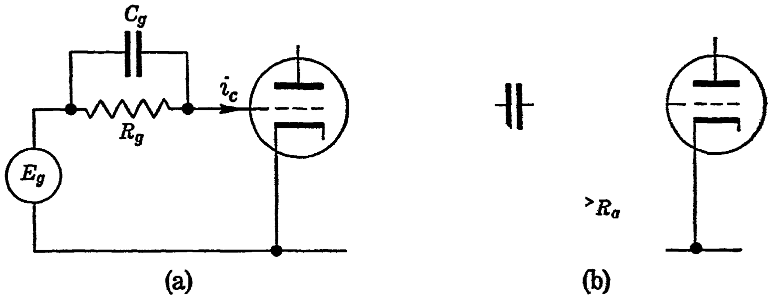


FIG. 11.14. Alternate methods of obtaining grid-leak bias in an amplifier which draws grid current.

consider the grid circuit as a half-wave rectifier with resistance load R_g shunted by smoothing capacitor C_g. The rectified and smoothed direct voltage developed across R_g approaches the peak value of the alternating grid voltage and provides a bias voltage of the correct polarity and magnitude to allow the grid-cathode voltage to become positive for a short period at the peak of each cycle. Compared with fixed bias from a battery or rectifier, this arrangement has the advantage of being self-adjusting. The bias automatically compensates for small changes in the alternating grid voltage and keeps the circuit operating at top efficiency.

Figure 11.14*b* shows another grid-bias arrangement. In this circuit C_g blocks the d-c component of grid current and forces it through R_g. Without inductor L_g the full alternating grid voltage appears across R_g and both an alternating and a direct component of current pass through it to produce a power loss. However, L_g presents a high reactance to the flow of alternating current without affecting the d-c component and thus reduces the power loss in the resistor.

Grid-leak bias has the definite disadvantage of allowing the

plate current to rise to a destructively large value should the input excitation fail. This danger can be eliminated by arranging a relay to remove the plate supply voltage when the input signal fails.

PROBLEMS

11.1 Curve 5 of Fig. 11.2 shows the Q of a coil consisting of 40 turns of No. 16 copper wire on a 6-in.-diameter form. (*a*) Compute the radio-frequency resistance at a frequency of 1 megacycle. (*b*) Compare this to the d-c resistance of the coil.

11.2 A series resonant circuit with a Q of 200 has an impedance of 10 ohms at the 1,000-kc resonant frequency. For a constant applied voltage of 0.1 volt, compute E_c at resonance and at 950 kc.

11.3 A pentode tuned amplifier is designed to provide a gain of 100 at a frequency of 2 megacycles. The coil Q is 200, and g_m equals 1,500 μmhos. Determine the values of L and C required for the resonant circuit.

11.4 A class C amplifier operates with a plate supply voltage of 2,000 volts, an alternating plate voltage of 1,250 volts, and a peak instantaneous current of 0.6 amp. Compute the maximum instantaneous amplifier efficiency.

11.5 The amplifier of Prob. 11.4 operates with an average plate current of 0.15 amp and a resonant load impedance of 7,000 ohms. Compute the power output, the power input, and the average plate efficiency of the tube.

11.6 The parallel resonant tank of the amplifier of Prob. 11.5 has an over-all Q of 15 which is much lower than the coil Q of 200 because of the resistance reflected into the circuit by the load. Compute (*a*) the required coil reactance, (*b*) the actual coil resistance, and (*c*) the total effective resistance. (*d*) From these resistances, compute the tank efficiency. Tank efficiency is the ratio between the power delivered to the reflected resistance as compared with the total a-c input from the tube.

11.7 A class C amplifier with the grid-leak bias arrangement of Fig. 11.14*a* requires a bias of 200 volts and draws an average grid current of 25 ma. To provide a sufficiently smooth bias the time constant of the bias circuit must exceed 10 cycles. Compute R_g and C_g for a frequency of 5 megacycles.

CHAPTER 12

OSCILLATORS

AN OSCILLATOR is an electronic device for producing an alternating power output without the use of mechanical moving parts. In a sense an oscillator acts as a converter of electrical energy from d-c form into an alternating output of controllable frequency. The electronic circuit accomplishing this result usually consists of an amplifier with positive feedback, a sort of self-excited amplifier that sacrifices part of the power output to produce its own input.

Oscillators perform many functions. They may be designed to produce alternating voltages of any frequency from a fraction of a cycle per second up to many billions of cycles per second, and a large number of different circuits have been evolved to cover this wide frequency range. Some, designed specifically for efficient power conversion, provide the high-frequency energy used for the induction heating of metals and the dielectric heating of poor conductors. These power oscillators usually sacrifice frequency stability and perfection of wave form in the interests of flexibility and efficiency. Other types of circuits provide especially good frequency stability. Each radio transmitter contains a crystal-controlled oscillator maintained at constant temperature for the sole purpose of producing an output frequency accurate to within about 10 parts per million. On the other hand, flexible laboratory oscillators controlled by a simple dial provide any desired frequency in the audio and supersonic bands with a frequency accuracy of about 1 percent and with a sinusoidal wave form containing only a few tenths of 1 percent distortion.

12.1 Power Oscillators

A power oscillator consists essentially of a class C tuned amplifier with a portion of the output voltage returned to the input as positive feedback. This arrangement produces a simple circuit

that works extremely well despite the fact that a complete quantitative analysis of its operation is practically impossible.

To understand the feedback arrangement in an oscillator of this type, we shall investigate the circuit of Fig. 12.1, which shows a

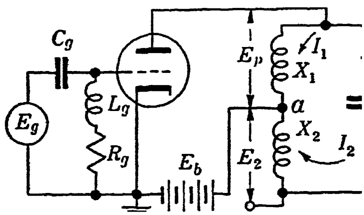


FIG. 12.1. A class C amplifier with a tapped resonant load.

conventional class C amplifier with grid-leak bias but with a tuned plate circuit (commonly called the tank) tapped at some point along the inductance. From the standpoint of the tube this circuit remains in resonance regardless of the position of tap a. To show this we shall first imagine a to be at the bottom of the coil. The circuit now becomes the normal parallel resonant circuit for which the total inductive reactance closely equals the capacitive reactance.

$$X_1 + X_2 = X_c \tag{12.1}$$

With tap a at the position shown on the diagram, the circuit consists of an inductive reactance X_1 in parallel with a capacitive reactance $X_c - X_2$ since X_c exceeds X_2 in magnitude. For resonance the parallel reactances must be equal.

$$X_1 = X_c - X_2 \tag{12.2}$$

But this is identical to Eq. (12.1); so the circuit is resonant for both connections, and the class C amplifier operates in much the same fashion as with a normal resonant load.

The purpose of tapping the circuit at point a is to provide an alternating voltage E_2 suitable for replacing the input voltage E_g. Since a single-stage amplifier produces a phase shift of 180 degrees, the voltage fed back to the input must be shifted another 180 degrees to obtain positive feedback. The vector diagram of Fig. 12.2 shows that E_2 is indeed nearly 180 degrees out of phase with E_p. Current I_2 leads E_p by an angle θ that approaches 90 degrees but does not quite reach it because of resistance in X_2. The voltage across X_2 likewise leads the current through it by an angle ϕ

that approaches a right angle. Consequently, voltage E_2 leads E_p by nearly 180 degrees. This analysis neglects the effect of mutual inductance between X_1 and X_2. Mutual inductance improves

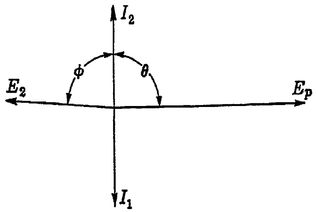


FIG. 12.2. A vector diagram showing that E_2 is nearly 180 degrees out of phase with E_p.

matters, but the circuit will work perfectly well with X_1 and X_2 constructed as separate coils.

Hartley Oscillator Circuit. Figure 12.3 shows a slight rearrangement of Fig. 12.1, but with E_2 now serving as the grid voltage.

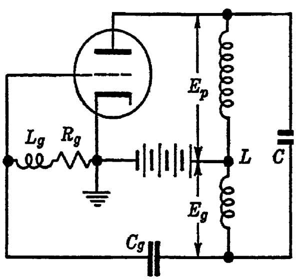


FIG. 12.3. Series-fed Hartley oscillator circuit.

This arrangement is called a Hartley oscillator. Capacitor C_g blocks off the d-c plate voltage while allowing the alternating component to pass to the grid. Resistor R_g together with L_g provides bias in the manner discussed in Sec. 11.4.

The oscillations start immediately upon application of the d-c plate voltage. Initially, the bias is zero because grid current must flow to produce a voltage drop in R_g. Application of the plate voltage therefore causes a sudden flow of plate current which excites the resonant circuit into oscillations. The alternating voltage E_g then produces an alternating component of plate current that maintains the oscillations of the resonant circuit, which in

turn produce the requisite input, etc. The first feeble oscillations produce only a small grid current and correspondingly low bias voltage so that the tube works in the class A region above cutoff. Under these conditions the feedback is larger than necessary, and the unstable oscillations increase in magnitude until limited by the increased grid bias. This shifts the operation down into the class C region and effectively reduces the net amplification by reducing the period of plate-current flow. The output stabilizes at the point where the circuit provides just sufficient output voltage to maintain the necessary input ($A\beta = +1$). The tuned circuit controls the frequency of oscillations, and the operating frequency nearly equals the natural resonant frequency.

With this class C operation the circuit converts the d-c input power into an alternating output with an efficiency approaching 70 percent. This is less than the efficiency of a good class C amplifier because the circuit must provide its own grid driving power.

The diagram of Fig. 12.4 shows a slightly different version of the

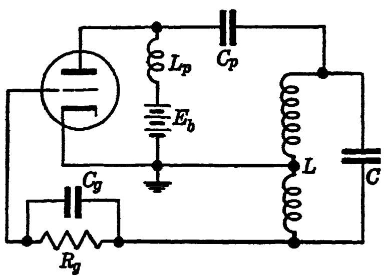


FIG. 12.4. Shunt-fed Hartley oscillator circuit.

Hartley circuit. This is called shunt feed because the direct component of the plate current does not pass through the tank circuit as it does in the series-fed circuit of Fig. 12.3. In the shunt-fed circuit the large inductive reactance of L_p forces the alternating components of plate current to pass to the tank through blocking capacitor C_p. This removes the direct voltage from the tank circuit, which may be an important consideration for a power oscillator operating from a plate supply of several thousand volts.

Colpitts Oscillator. The Colpitts circuit of Fig. 12.5 shows another basic type of oscillator similar to the Hartley but with the tuned circuit tapped by splitting the capacitor into two sections.

This provides a feedback voltage of proper phase in almost the same fashion as does the tapped inductor of the Hartley oscillator.

The Colpitts circuit has the disadvantage that two adjustable

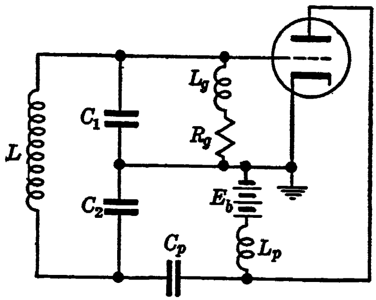


FIG. 12.5. Colpitts oscillator.

capacitors are required for adjusting the resonant frequency. For this reason most oscillators designed to cover a range of frequencies employ variations of the Hartley circuit. In other respects there is little to choose between the two arrangements.

12.2 High-frequency Heating

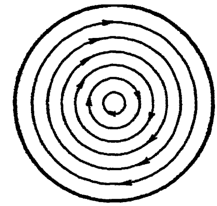
An increasingly important application of power oscillators is the generation of power for radio-frequency heating. While this is an expensive method of obtaining heat, it offers advantages of speed and control that make it worth while for many specialized jobs. For example, the production of tin plate with electroplating, followed by induction heating to momentarily melt the tin and close the surface pores, produces a satisfactory coating with only one-third of the tin required by the old hot-dipping process.

There are two different types of radio-frequency heating: induction heating and dielectric heating. Induction heating must be employed to heat electrically conducting materials because it operates on the eddy-current principle. Dielectric heating provides a method of heating nonconductors by means of the power loss in such materials when subjected to intense alternating electric fields.

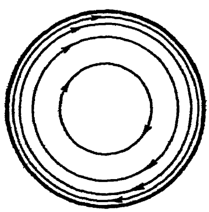
Induction Heating. The useful frequency range for induction heating extends from less than 5 kilocycles to several hundred kilocycles, depending upon the application. The lower frequencies are used for heating large metallic masses in an induction furnace

where the whole volume must be brought to a high temperature. Rotating machinery often supplies the relatively large power requirements of such installations. Power for the higher frequency installations, however, comes from vacuum-tube oscillators.

An advantage of using higher frequencies, as suggested by Fig. 12.6, is that the high eddy-current density near the outer surface confines the majority of the heating to the surface. This rapid surface heating followed by quenching makes it possible to produce quickly a casehardening that would require much longer treatment by conventional methods. Proper frequency choice for a given shaped piece of work controls the hardening depth, and the whole process can be carried out continuously by passing the parts on a moving conveyer through the high-frequency coil. Convenience, speed, and relatively low heat waste make the process economically feasible despite the high cost of heating per unit of energy.



(a) Relatively uniform eddy-current flow at low frequencies



(b) Surface concentration of eddy-current flow at higher frequencies

FIG. 12.6. Eddy-current flow in a round bar which is placed in a coil carrying alternating current.

Induction heating also does a particularly neat job of brazing small parts. The assembled parts, treated with flux and with a length of brazing wire fashioned to fit closely to the areas where brazing is desired, are placed in an induction heating coil with power applied for a predetermined period just sufficient to melt the wire and join the metal pieces. The absence of flame makes it easy to process the work in a reducing atmosphere if required, and the resulting clean product needs little further work to make it presentable.

The Colpitts oscillator can be adapted to the production of heating power by making the heating coil part of the total circuit inductance, as shown by Fig. 12.7. Since this coil carries the heavy tank current (about Q times the alternating plate current), it produces a strong alternating magnetic field and rapidly heats the work. Of course, placing the metal work inside the heating coil

does change the inductance and affects the oscillator frequency somewhat. With a Hartley circuit, however, the work coil would necessarily have to be placed to one side of the inductance tap;

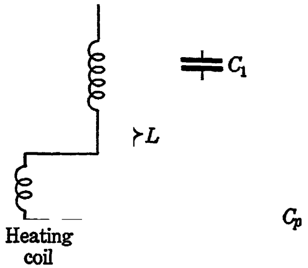


FIG. 12.7. An adaption of the Colpitts oscillator tank circuit to induction heating.

placing the work in the coil would change the balance between the two portions of the inductance and seriously affect the circuit operation.

Dielectric Heating. Dielectric heating refers to the heating produced in nonconductors (dielectrics) subjected to a strong alternating electric field. The molecules of insulating materials are more or less polar; that is, the centroids of the positive and negative charges have slightly different positions which, from the electrical standpoint, make each molecule equivalent to one positive and one negative charge separated by a flexible lever arm. In a steady electric field the forces on the charges produce a turning moment tending to align the molecule with the field and actually stretch it, but this is opposed by the thermal agitation which continually upsets the orientation. With an alternating electric field the molecules oscillate rapidly, their force reactions with one another add to the general confused motion, and the energy lost produces a temperature rise.

As with induction heating, dielectric heating produces the heat in the material itself, which makes it possible to raise the temperature rapidly. To obtain dielectric heating the work is placed between a pair of metal plates connected to a high-frequency oscillator capable of supplying a high voltage. Figure 12.8 shows a typical arrangement, with a sandwich of two wood slabs and glue placed between a pair of parallel electrodes for the purpose of heating and setting the glue. This problem is essentially that en-

countered in the production of plywood. With externally applied heat it is impossible to heat the glue line rapidly without scorching the outer wood surface because all the heat has to pass through

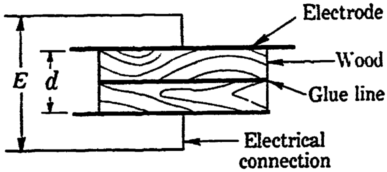


FIG. 12.8. Dielectric heating of wood to set a glued joint.

the wood to reach the center. With dielectric heating, however, heat generated directly in the wood and glue heats the center rapidly while the heat removed from the outer surfaces keeps them relatively cooler. This greatly reduces the time required to set the glue and permits dielectric heating to compete with less expensive forms of heat energy.

An analysis of the energy loss in a uniform dielectric between two parallel electrodes shows that the heat produced per unit volume is

$$H = \frac{(0.55 \times 10^{-12})(f)(E^2)(\epsilon)(pf)}{d^2} \quad \text{watts per cc} \tag{12.3}$$

In this equation f is the frequency, E the voltage between plates, d the plate spacing in centimeters, ϵ the dielectric constant, and pf represents the power factor of the material. Here, as elsewhere, the power factor is the cosine of the phase angle between the applied alternating voltage and the current flow through the material. To produce sufficient heating with a voltage below the sparking potential requires the use of frequencies well above 1 megacycle per second. At 13 megacycles (one of the frequencies reserved for dielectric heating by the Federal Communications Commission) a typical heating application will require an applied voltage of several thousand volts.

12.3 Oscillators for Frequency Control

Ordinary power oscillators are not particularly stable devices; the frequency of oscillation approximately equals the natural frequency of the tuned circuit, but load changes, replacement of tubes, and d-c supply voltage variations all affect the frequency to some extent. In addition to this the resonant circuit itself ex-

pands when heated by the power losses. This affects the inductance and capacitance and makes the frequency dependent upon the temperature.

For these reasons the frequency of a large power oscillator may change several percent when the load is applied, but a small unloaded oscillator of the same type with special attention paid to the mechanical stability of the parts may hold the frequency to a few tenths of 1 percent. While this may seem good compared with the ordinary accuracy of voltage and current measurement, much better frequency control than this can be obtained.

The need for excellent frequency control extends to both communications and timing. In the broadcast band, for example, the stations occupy a frequency space about 1 percent wide; if they are not to transgress into the next channel, the frequency must be held to a tiny fraction of 1 percent. At higher frequencies the number of communications carried by a given frequency band depends essentially upon the accuracy with which each frequency can be maintained by each transmitter and selected by the receivers. Timing also requires precision frequency control. The best standards of time are now crystal-controlled oscillators; for example, station WWV operated by the U.S. Bureau of Standards continuously broadcasts frequencies accurate to within 1 part in 10 million.

The frequency control in these precision oscillators comes from a vibrating mechanical element, rather than from a tuned circuit, because a mechanical oscillator provides much stiffer control; in effect it has a much higher Q. A comparison of mechanical and electrical resonance curves shows that a well-designed mechanical oscillator (a tuning fork, for example) may possess an equivalent Q of over 10,000. In fact a quartz crystal mounted in vacuum to reduce the air damping may reach a Q of several hundred thousand.

Of course these mechanical elements change their physical dimensions with the temperature, and it is necessary either to make them temperature compensated or to operate them in a constant-temperature chamber. The most precise frequency standards do both.

Figure 12.9 shows the circuit diagram for a simple crystal oscillator. The tuned plate circuit does not control the frequency but merely provides a conveniently adjustable load impedance.

Feedback takes place through the grid-plate capacitance (at radio frequencies), and L_g and R_g provide the bias when the circuit oscillates and draws grid current. The circuit operation depends

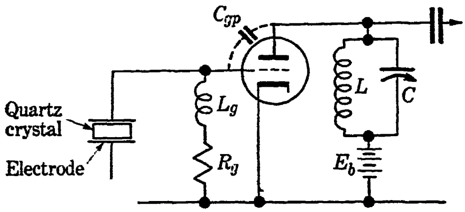


FIG. 12.9. A simple crystal oscillator circuit.

upon the *piezoelectric* property of quartz. Many other crystals also exhibit this effect.

The term piezoelectric refers to the generation of an electromotive force between the faces of a crystal when pressure is applied. Conversely, application of voltage to the crystal faces produces a corresponding mechanical distortion. Placing metal electrodes on the crystal face makes it possible to obtain an electromechanical coupling between the vacuum-tube circuit and the crystal itself. An alternating voltage applied to the electrodes produces mechanical vibrations of the quartz, and at the resonant frequency the vibrations are particularly strong. From the electrical standpoint the crystal-electrode assembly appears as a sharply resonant circuit capable of controlling the frequency of oscillations.

Quartz makes a particularly good resonator because of its chemical and mechanical stability. It is hard and reasonably strong, and its temperature coefficient of expansion is rather low; in fact, suitable orientation of the crystal dimensions with respect to the crystalline structure produces a resonant frequency independent of the temperature over a reasonable range. This stability, together with the high equivalent Q provides a frequency accurate to within 100 parts per million without difficulty. Additional refinements improve the stability.

The crystal size varies with the frequency. For low frequencies of about 100 kilocycles the crystal is often in the form of a rectangular bar an inch or two long. At higher frequencies a thin quartz slice about the size of a postage stamp serves as the resonant element. Above about 30 megacycles the crystal becomes too thin to withstand the necessary vibrations, and harmonics must

be employed. Stable fixed audio-frequency oscillators usually employ tuning forks because the necessary size of quartz bar becomes too great.

12.4 Wide-range Oscillator

In contrast to the precise single-frequency crystal oscillator, the wide-range oscillator provides, at the turn of a dial, any desired frequency within a broad band. A necessary piece of equipment in any properly equipped laboratory, this device finds its use in testing the frequency response of amplifiers, for measuring unknown frequencies by comparison methods, and as a low-power source of alternating voltage.

Radio-frequency oscillators of this type commonly employ a simple Hartley circuit with a variable capacitor to cover a frequency range of about 3.5 to 1 and a band switch similar to that of radio receivers to change the value of the inductance in steps. Such signal generators, as they are called, greatly aid the testing and servicing of radio equipment.

Tuned-circuit oscillators for the audio-frequency band require large values of inductance and capacitance which are bulky and not conveniently adjustable. Consequently, other types of oscillator circuits have been developed. One of the best of these is the resistance-capacitance, or R–C, oscillator of Fig. 12.10.

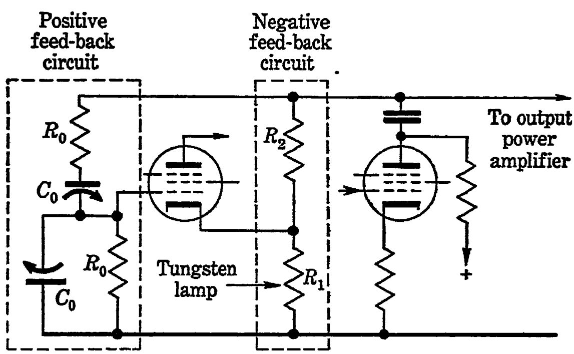


FIG. 12.10. Functional diagram of an R–C oscillator to show the two feedback loops. The actual amplifier details have been left out.

This circuit has been made to cover frequencies from below 1 cycle per second up to several megacycles. A typical commercial model for the audio and supersonic bands covers from 20 to

200,000 cycles in four decades controlled by a single dial and band switch. The frequency stability is good, and the accuracy of setting the frequency depends to a large extent upon the mechanical stability of the variable tuning capacitors; 1 percent is a common figure. The oscillator also produces an especially pure wave form with less than 0.5 percent of harmonic distortion.

The R–C oscillator consists of a two-stage resistance-capacitance-coupled amplifier with two feedback circuits. The diagram of Fig. 12.10 omits the familiar amplifier details to concentrate attention on the feedback arrangement. Resistors R_1 and R_2 provide negative feedback to the cathode of the input tube; since no reactances are included, the negative feedback is independent of the frequency, as shown by Fig. 12.11. The positive feedback

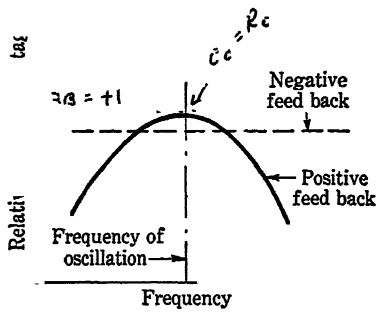


FIG. 12.11. At the frequency of oscillation the positive feedback exceeds the negative by just enough to make $A\beta = +1$.

loop to the input grid consists of two equal fixed resistors (each frequency band requires a different value of resistance) and a two-section variable condenser which control the output frequency. This arrangement provides a feedback voltage that reaches a maximum at some middle frequency and drops off above and below that point, as shown by Fig. 12.11.

The reason for this curve shape is not particularly difficult to understand. At low frequencies the reactance of C_0 becomes large compared with R_0, and the series C_0–R_0 combination has an impedance very large compared with the parallel C_0–R_0 pair. Thus the fraction of the output voltage that appears across the parallel combination becomes small. At high frequencies the reactance of C_0 becomes small compared with R_0, the series impedance approaches R_0, the parallel impedance approaches the re-

actance of C_0, and again the fraction of the output voltage that reaches the input grid is small. The maximum point between these extremes happens to occur at the frequency for which the reactance of C_0 equals R_0.

To produce oscillations the negative feedback is adjusted so that the *peak* positive feedback just exceeds the negative feedback by enough to make the product $A\beta$ equal to $+1$. The adjustment is critical; with too much negative feedback the circuit fails to oscillate, and with slightly too much positive feedback the oscillations increase in magnitude until the wave form becomes distorted. In fact, the success of this circuit depends upon the choice of a small tungsten lamp for resistor R_1. This lamp has a low cold resistance which increases rapidly when sufficient current passes through the filament to heat it. In the oscillator circuit the current through R_1 consists of the small plate current of the input tube plus an alternating component flowing through R_2. The size of this alternating component, and consequently its heating effect on R_1, depends on the magnitude of the output voltage. Thus with feeble oscillations the net current through R_1 is low, the negative feedback is also low, and the excess positive feedback causes the oscillations to build up. This simultaneously increases the value of R_1 by heating and increases the negative feedback until the output becomes stable.

This clever arrangement provides an exceptionally constant output voltage and good frequency stability despite wide variations in the d-c supply voltage. An extension of the principle to very low-frequency oscillators of 1 cycle per second has been successful, but difficulty is usually encountered with the tungsten lamp. At these frequencies the thermal lag of the lamp cannot average out the heating effect over the period of any one cycle, and the lamp resistance tends to follow the individual alternations.

CHAPTER 13

AMPLITUDE MODULATION AND DEMODULATION

IN ELECTRICAL-circuit theory the term modulation applies to the process of impressing the characteristics of one wave upon another. Speech provides a familiar physical example in which the mouth modulates a sound stream to produce recognizable words. Merely producing a steady tone conveys no intelligence. Likewise, the transmission of a steady signal by a radio-frequency transmitter tells nothing except that the station is or is not operating, but useful information can be sent by turning it on and off according to some prearranged code. This simple form of *amplitude* modulation is the basis for the familiar dot-dash code. Speech, music, and other wave forms can be transmitted by a refinement of this system which varies the amplitude of the transmitted wave in direct proportion to the sound or picture signal. The high frequency carrying this modulation is called the *carrier*.

A *modulator* is a device that receives the incoming streams of carrier wave and signal and combines them to produce a modulated output. There are a large number of different circuits and devices for accomplishing this result, but a fundamental property of electrical circuits used to produce modulation is nonlinearity. In fact, because all vacuum tubes are nonlinear, undesired modulation often occurs when not expected.

Demodulation is the inverse process which removes the modulating information from the carrier to recover the original signal. All radio receivers, for example, contain demodulators to recover the original sound signals from the modulated radio-frequency wave. Likewise many electromechanical measuring systems employ modulated carriers to transmit information from the test point to a central recording unit. Here demodulators remove the modulation from the carrier before recording.

13.1 Simple Amplitude-modulated Wave

For amplitude modulation the amplitude of the carrier wave is varied in direct proportion to the intelligence transmitted. Figure 13.1 shows such a wave with simple sinusoidal modulation;

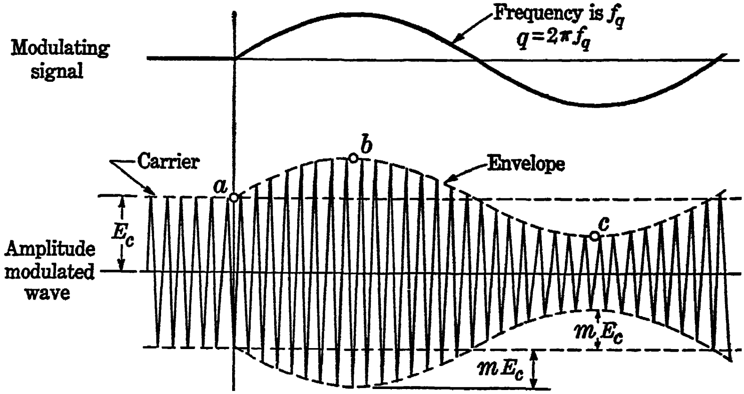


FIG. 13.1. Simple amplitude-modulated wave with sinusoidal modulation.

the upper curve represents the modulating signal, and the lower one shows the resulting amplitude-modulated wave. The dotted *envelope abc*, which outlines the amplitude of the modulated wave, accurately follows the modulating signal. Up to point a no modulation takes place and the carrier peaks are constant in amplitude. At point b the modulated wave reaches the *crest*, and at point c it falls to a minimum called the *trough*. Half the vertical distance between crest and trough divided by the unmodulated amplitude E_c is called the modulation factor m, often expressed in percent. The diagram shows about 50 percent modulation. A value of 100 percent represents the maximum possible with conventional circuits. For this limiting case point c falls on the zero line.

The equation of envelope abc is $E_c + mE_c \sin qt$, where q for the modulating signal corresponds to ω for the carrier frequency. This envelope represents the peak amplitudes of the individual carrier-frequency waves. The equation of an individual carrier wave is $e = E_{\max} \sin \omega t$. Replacing $E_{\max}$ by the expression for the envelope, we obtain

$$\begin{aligned} e &= (E_c + mE_c \sin qt)(\sin \omega t) \\ &= E_c \sin \omega t + mE_c \sin qt \sin \omega t \end{aligned}$$

With a trigonometric substitution for the product of two sines, we get

$$e = E_c \sin \omega t + \frac{mE_c}{2} \cos(\omega - q)t - \frac{mE_c}{2} \cos(\omega + q)t \quad (13.1)$$

This result shows that the modulated wave of Fig. 13.1 can be thought of as three separate sine waves of constant amplitude. The first of these is the unmodulated carrier, as shown to the left of point a. The other two, identical in size, differ from the carrier frequency by plus and minus the modulating frequency. These are called the *side frequencies*. For example, a carrier of 1,000 kilocycles modulated with an audio frequency of 3000 cycles will contain frequencies of 997, 1,000, and 1,003 kilocycles.

When the modulation consists of speech or music with a changing complex wave form, the envelope has a corresponding shape and there are many simultaneous side frequencies varying in magnitude and frequency. Such a group is called a *side band.* The presence of side bands requires that radio receivers be designed with selective circuits that pass a band of frequencies rather than just the carrier frequency. Likewise, broadcast stations are separated in frequency by 10 kilocycles to prevent interference between their side bands. Figure 13.2 shows typical side-band

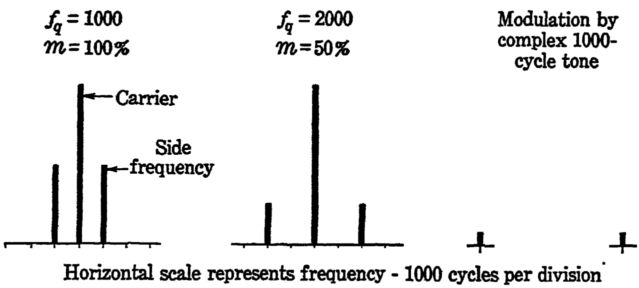


FIG. 13.2. Bar diagram showing the side-band patterns for several amplitude-modulated waves.

patterns for several different modulations. In each diagram the height of the vertical bar represents the side-band amplitude.

The relation between the side-frequency concept and the picture of Fig. 13.1 can nicely be shown with the aid of the vector diagram of Fig. 13.3. The three vectors correspond to a carrier rotating with an angular velocity ω and two side frequencies rotating with velocities $\omega + q$ and $\omega - q$. If, now, the observer imagines himself

to be rotating with an angular velocity ω as well, the carrier vector will appear to stand still but the two side-frequency vectors will rotate in opposite directions with velocities $+q$ and $-q$. The

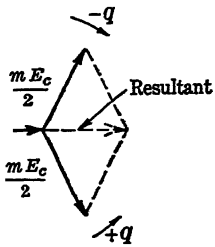


FIG. 13.3. Vector diagram of a simple amplitude-modulated wave.

phase positions of these side-frequency vectors are such that at every instant their perpendicular components cancel, leaving only the horizontal components. These add to or subtract from the length of the carrier vector without changing its phase position. Consequently, the sum of the three rotating vectors is equivalent to a single vector of sinusoidally varying length rotating at the carrier frequency. The vertical projection of such a vector would sweep out the modulated wave of Fig. 13.1.

13.2 Modulation in Nonlinear Circuits

Although the type of modulation about to be described is not always the most practical or efficient method, it is of basic theoretical and practical importance because it illustrates aspects of the modulation problem not displayed by Fig. 13.1 and Eq. (13.1). We shall study a nonlinear circuit having a curve of output voltage or current plotted against input voltage that takes the form of Fig. 13.4. The circuit to produce this might contain a

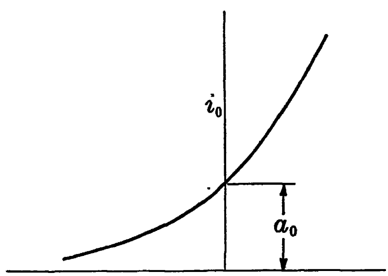


FIG. 13.4. Typical nonlinear characteristic.

diode, triode, or multigrid vacuum tube, a disk rectifier, or perhaps a nearly saturated iron-clad circuit. Figure 13.5 shows a typical

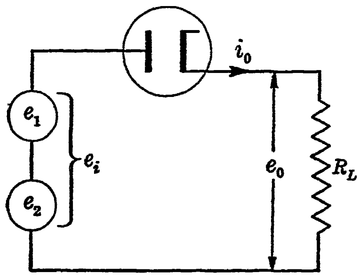


FIG. 13.5. Simple nonlinear circuit.

simple series circuit consisting of two sinusoidal input signals, a diode, and a load resistor to change current variations into useful voltage variations. The curve of Fig. 13.4 represents the dynamic characteristic for the *whole circuit*, not just that of the diode alone.

We shall imagine that two sine waves of different frequencies are simultaneously being applied to the circuit to find the resulting interaction between them in the output. The graphic picture of

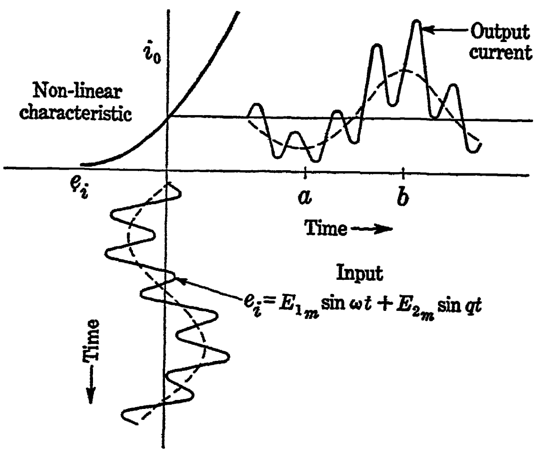


FIG. 13.6. A graphic analysis of the nonlinear circuit of Fig. 13.5.

Fig. 13.6 shows these waves as a low and a high frequency, their sum being the signal applied to input. The output wave, ob-

tained graphically, looks similar to the input, but it differs from it in several important respects. Careful inspection will show that at time b the high-frequency component of the complex wave has a greater amplitude than at time a. Thus the amplitude of the high-frequency component now varies at a rate corresponding to the low frequency, and modulation has been effected. Likewise, the individual high-frequency waves become distorted and contain harmonics and, although this is less apparent, the same holds true for the low-frequency component.

To investigate this output wave we shall attempt a more analytical study of the situation. A small portion of any smooth curve can be represented by a polynomial of relatively few terms. Since the curve of Fig. 13.4 is of this type, we can write

$$i_0 = a_0 + a_1 e_i + a_2 e_i^2 + a_3 e_i^3 + \cdots$$

Term a_0 represents the operating point current. Coefficient a_1 is the slope of the curve at the operating point, and if the curve were a straight line only the first two terms of the series would be required. For a smooth curve additional terms are necessary but, in general, the coefficients decrease rapidly as the order increases. In fact for our analysis let us assume that the curvature is so gentle as to require only the first three terms for reasonable accuracy. Thus

$$i_0 = a_0 + a_1 e_i + a_2 e_i^2 \tag{13.2}$$

This can always be made true for small signals using only a little of the curve. This point of view is different from that employed in Chap. 3 for the discussion of single-phase rectifier circuits. There with large voltage swings the rectifier was considered to be either a good conductor or an open circuit, depending upon the direction of current flow. Here the applied voltage is small, and only a small portion of the curve near zero is of interest.

Corresponding to the input wave shown in Fig. 13.6

$$e_i = E_{1m} \sin \omega t + E_{2m} \sin qt$$

where ω is $2\pi f_1$ and q is $2\pi f_2$, f_1 and f_2 being the two different input frequencies. Placing this voltage into Eq. (13.2), we obtain

$$\begin{aligned} i_0 = a_0 + a_1 E_{1m} \sin \omega t + a_1 E_{2m} \sin qt + a_2 E_{1m}^2 \sin^2 \omega t \\ + 2a_2 E_{1m} E_{2m} \sin \omega t \sin qt + a_2 E_{2m}^2 \sin^2 qt \end{aligned}$$

With the terms rearranged and trigonometric substitutions made for $\sin^2$ and $\sin \omega t \sin qt$, this expression becomes

$$
\begin{aligned}
i_0 = a_0 &+ \frac{a_2}{2}E_{1m}^2 + \frac{a_2}{2}E_{2m}^2 + a_1E_{1m}\sin \omega t + a_1E_{2m}\sin qt \\
&- \frac{a_2}{2}E_{1m}^2 \cos 2\omega t - \frac{a_2}{2}E_{2m} \cos 2qt \\
&+ a_2E_{1m}E_{2m}\cos(\omega - q)t - a_2E_{1m}E_{2m}\cos(\omega + q)t \qquad (13.3)
\end{aligned}
$$

The first three terms of Eq. (13.3) do not vary with time and represent the d-c component. With no input signal, voltages E_1 and E_2 are zero and i_0 becomes merely a_0. Therefore a_0 is the quiescent or operating point current, and the next two terms of the equation must represent the increase in average current caused by the application of the input signals. This same increase was observed in connection with triode amplifier distortion. The fourth and fifth terms represent the two input frequencies, unchanged except for amplitude. These are followed by two double-frequency components representing the second harmonic distortions of each original frequency, as might be expected. Last come two new frequencies, one equal to the sum and the other equal to the difference of the original input frequencies. These are called the beat, or heterodyne, frequencies, and they are the ones of greatest importance in modulation or demodulation.

It is essential to understand that these new sum and difference frequencies do not exist until the two input voltages have been applied to a nonlinear device. Simply adding two waves of different frequencies does not of itself produce the new beat frequencies. It is true that the ear can actually hear a beat between two sound waves of nearly the same frequency, but this consists of the changing sound intensity as the two component sound waves alternately come into and then out of phase. No true sound wave having the difference frequency appears until a nonlinear element has been encountered.

If the curve contains higher order terms than the second, additional third- and fourth-order terms appear in Eq. (13.3), such as 3ω, $\omega + 2q$, $2q + \omega$, etc., but the original sum and difference frequencies are not affected.

13.3 Production of a Modulated Wave

Square-law Modulation. The complicated expression of Eq. (13.3) contains, among other things, three terms that correspond

to the three components of the simple amplitude-modulated wave of Fig. 13.1. This fact suggests using a resonant circuit to discard all but the three desired components and produce an amplitude-modulated wave. Figure 13.7 shows such an arrangement with a

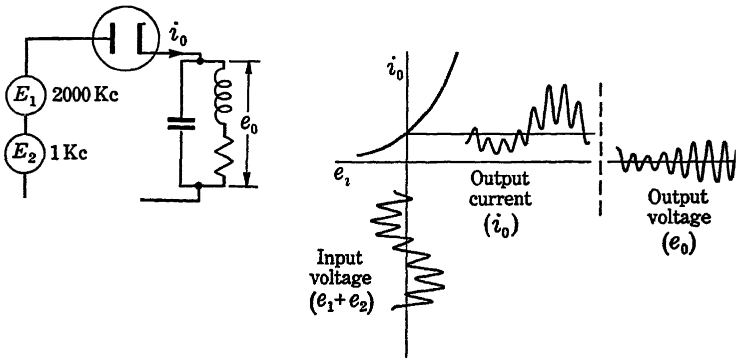


FIG. 13.7. The production of an amplitude-modulated wave by a nonlinear circuit.

resonant load impedance tuned to the carrier frequency. Voltage E_1 represents a constant-amplitude sine wave of the carrier frequency, say 2,000 kilocycles. The modulating sine wave is E_2, 1,000 cycles, for example. The circuit current i_0 contains all the frequency components given by Eq. (13.3), but the tuned load has an appreciable load impedance only in the vicinity of the resonant frequency. Thus the carrier frequency (2,000 kilocycles) and the two sum and difference frequencies (1,999 and 2,001 kilocycles) are the only ones to appear in the output voltage. With a sufficiently broad resonant circuit to present essentially the same impedance R to these frequencies, the output voltage produced by i_0 becomes

$$e_0 = a_1RE_{1m} \sin \omega t + a_2RE_{1m}E_{2m} \cos(\omega - q)t \\ - a_2RE_{1m}E_{2m} \cos(\omega + q)t \quad (13.4)$$

This is an equation of exactly the same form as Eq. (13.3), consisting of a carrier flanked by two side frequencies. Therefore the output wave looks exactly like the simple amplitude-modulated wave of Fig. 13.1. The graphic analysis of Fig. 13.7 shows the same result.

From a practical standpoint this type of circuit is limited to small percentages of modulation. Greater modulations require

operating over so much of the curve that the circuit no longer behaves in a square-law fashion; that is, additional terms above the second order must be included in the analysis. Under these conditions other side bands occur, and the modulation envelope becomes distorted.

Linear Modulation. The term linear modulation refers to a system in which the amplitude of the modulated envelope is directly proportional to the instantaneous value of the modulating signal. It does not refer to the internal characteristics of the circuit; in fact, a perfectly square-law circuit produces linear modulation because the output wave, as given by Eq. (13.4), exactly follows the sinusoidal modulating voltage e_2. The only reason that a square-law modulator fails to give linear modulation at high modulation levels is that practical nonlinear devices do not remain square law over a wide operating range.

The linearity of an ideal modulator can be expressed as a graph (Fig. 13.8), showing a straight-line relationship between the instantaneous modulating voltage e_q and the crest (or effective) carrier output E_0. There are many types of circuits which give linear modulation. One common type employs a tuned class C amplifier with the modulating voltage applied in series with the d-c plate supply. Since an efficient class C amplifier operates with a peak alternating output voltage that closely approaches the d-c supply voltage, varying this supply voltage causes the alternating output to vary in nearly direct proportion. This arrangement produces one of the most linear and efficient modulators known; for this reason many radio transmitters employ modulated class C amplifiers.

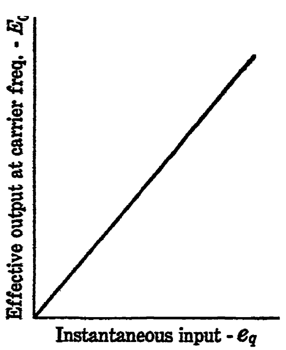


FIG. 13.8. Characteristic curve of an ideal linear modulator.

Modulation is not necessarily a purely electrical process. Chapter 15 describes an electromechanical arrangement which produces a modulated carrier in response to mechanical input and thus transforms a mechanical displacement into an electrical signal. In an arrangement of this type, the input signal may have no frequency components above 50 cycles and the carrier frequency can be as low as 1 kilocycle.

13.4 Frequency Shifting—The Superheterodyne

An important aspect of modulation in nonlinear circuits is the property of frequency shifting. The last two terms of Eq. (13.3), which are exactly alike except that one represents the difference and the other the sum of the two input frequencies, exhibit this property. Furthermore, the amplitudes of these two components are directly proportional to the *product* of the two input signal amplitudes. This makes it possible to combine a modulated input signal with another voltage of fixed frequency and amplitude to produce an output of new frequency but which faithfully follows the amplitude (and perhaps frequency) variations of the original. Of course, the circuit output contains many other frequency components as well, but these can be eliminated with the aid of tuned circuits.

An important application of this principle is in the superheterodyne type of radio receiver which takes advantage of the fact that amplification and selection of the lower radio frequencies is simpler than that of the higher ones. The block diagram of Fig. 13.9 illustrates this process.

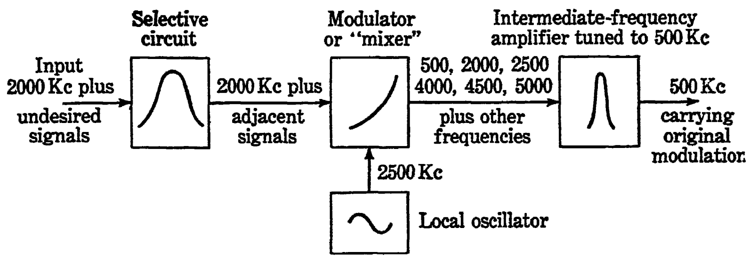


FIG. 13.9. A block diagram showing the operation of a superheterodyne receiver.

We shall imagine the incoming signal to consist of a desired 2,000-kilocycle signal plus a number of other unwanted frequencies. The first selective circuit effectively eliminates those frequencies considerably different from 2,000 kilocycles. A simple resonant circuit can do this. The modulator input then consists of the desired signal plus others of nearby frequencies. After combination with the local oscillator frequency of 2,500 kilocycles, the modulated mixture passes through a sharply tuned amplifier adjusted to 500 kilocycles. This intermediate-frequency amplifier rejects all frequencies except the difference frequency between the desired signal and the local oscillator. Since the amplitude of this new

frequency varies in direct proportion to the original input, the intermediate frequency accurately reproduces the desired input and carries the same information.

The advantages of this system over straight amplification of the original signal are that (1) it is easier to build high-gain selective circuits at the lower radio frequencies; (2) two signals of, say, 2,000 and 2,020 kilocycles are hard to separate, but the difference frequencies of 500 and 480 kilocycles have an effective separation four times as great expressed in percent; (3) the number of variable tuned circuits needed for a receiver designed to cover a range of frequencies is limited to the oscillator and input selective circuits; the intermediate-frequency amplifier provides the majority of the selectivity with *fixed* tuned circuits that are cheaper and easier to keep in adjustment.

13.5 Demodulation

After the modulated carrier has been transmitted as a radio wave or by wire, and amplified by the receiver, it must be demodulated to recover the original modulating signal. The most common circuit for doing this is the simple half-wave rectifier with smoothing condenser, discussed in Chap. 3. Figure 13.10

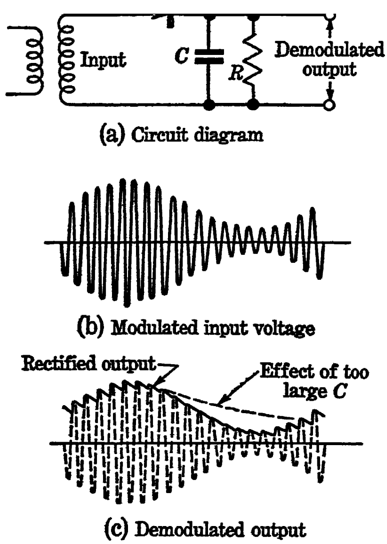


FIG. 13.10. Circuit diagram and wave forms showing the operation of a simple diode envelope demodulator.

illustrates its application to the process of demodulation. The rectifying element may be either a crystal rectifier or a vacuum diode. Load resistor R_L is made large, usually over 100,000 ohms, to obtain maximum output voltage with minimum input power. Capacitor C serves to smooth the rectified output by filling in the gaps between individual waves, as shown by the heavy output-voltage line. This gives a rough approximation to the envelope of the input voltage; for this reason the circuit is called an envelope demodulator.

To illustrate the details of the process better, Fig. 13.10 shows a carrier frequency only 20 times the envelope frequency. In actual practice the frequency ratio may be several hundred or more, and the saw-tooth output closely approaches the envelope. Capacitor C must be carefully chosen to provide as much smoothing as possible without filling in the envelope troughs, as suggested by the dotted line. A full-wave circuit simplifies this task by providing twice as many peaks and a correspondingly closer approximation to the envelope.

The graphic analysis of Fig. 13.10 assumes that the rectifier conducts perfectly in the forward direction with negligible voltage drop. This approximation holds so long as the input voltage remains higher than about 1 volt for typical diode circuits. For smaller signals we must analyze the circuit from the square-law standpoint to find that demodulation does take place, but with considerable distortion. Since it is easy to provide a large amplified voltage for the demodulator, an analysis of nonlinear demodulation becomes of secondary importance.

PROBLEMS

13.1 The modulation pattern displayed on a cathode-ray oscilloscope looks similar to Fig. 13.1, except that the individual carrier sine waves cannot be distinguished. Measurement of the envelope pattern shows a maximum width of 3 in. and a minimum of 1.2 in. Compute the percent modulation.

13.2 The 2-megacycle carrier of a radio transmitter has an effective value of 100 volts, and it is 60 percent modulated with a sinusoidal wave having a frequency of 3,000 cps. (*a*) Write the envelope equation. (*b*) Compute the magnitudes of the component frequencies in the modulated wave, and draw a bar diagram showing these components.

13.3 A pure sine wave of 0.5 volts applied to a nonlinear circuit produces an output current with an average value of 5 ma, a fundamental component of 2 ma effective value, and a second harmonic of 0.15 ma effective. (*a*) Compute the constants a_0, a_1, and a_2 for the circuit. (*b*) What would be the current without an input signal?

13.4 The output of a strain-gage bridge contains a 2,000-cycle carrier modulated with a complex wave shape containing frequencies up to 100 cycles. This wave then passes through a 2,000-cycle tuned amplifier which employs a resonant circuit with an effective Q of 25. Is the amplifier suitable for amplifying the bridge output? Explain.

13.5 The diode demodulator of Fig. 13.10 demodulates a carrier of 500 kc carrying modulation frequencies up to 5 kc. The load resistance of 200,000 ohms is shunted by a capacitance of 0.002 μf. Criticize the choice of shunt capacitance.

13.6 The output of an electromechanical pickup consists of a 2,000-cycle carrier modulated 30 percent at a frequency of 5 cps. This voltage must be first amplified and then demodulated to produce a 5-cycle output having an effective value of 10 volts. Compute the approximate amplification required to obtain this output with an input carrier of 2 mv effective value.

CHAPTER 14

THE CATHODE-RAY OSCILLOSCOPE

FOR MANY years physicists and engineers have employed mechanical systems for tracing out wave forms and curves showing the operation of their apparatus. These contrivances work moderately well at low frequencies, but when the vibration frequency exceeds several thousand cycles per second, even the cleverest and most carefully designed mechanical apparatus fails because of the terrific accelerations required of the moving parts. And at radio frequencies the situation is hopeless.

The solution to this problem was the gradual development of the cathode-ray oscilloscope. The term cathode ray, a holdover from the early physical experiments, refers to an electron stream. Electrons possess three distinct advantages: (1) they are the lightest charged particle known, (2) for their mass they possess a very large electrical charge, and (3) they are easily observed visually or photographically. The first property makes extremely rapid motion possible, the second permits simple control of the motion, and the third provides a visual interpretation that appeals to the senses.

14.1 The Cathode-ray Tube

The cathode-ray tube in its present-day form consists of three distinct parts, shown in Fig. 14.1. These are (1) an electron gun

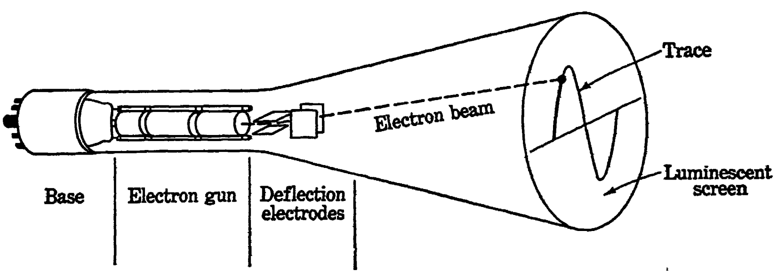


FIG. 14.1. Simplified diagram to show the essential parts of a cathode-ray tube.

for producing a concentrated beam of high-speed electrons, (2) a set of deflection electrodes for moving the beam as desired, and (3) a screen that produces light when struck by the electrons.

Electron Gun. The electron-gun assembly of Fig. 14.2 consists

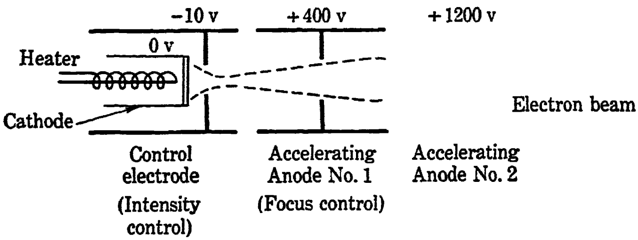


FIG. 14.2. Simple electron gun with electrostatic focusing.

of a series of metal cylinders arranged on a common axis. The electrons start from a small spot of oxide coating on the end of an indirectly heated cathode and then pass through a control electrode consisting of a closed cylinder with a small central hole. This further restricts the dimensions of the beam, and a negative potential applied to the electrode controls the beam intensity exactly as the grid bias controls the plate current of a triode. The control-electrode aperture, however, does not restrict the beam sufficiently; without additional electrodes the electrons would only spray out in a generally forward direction.

The two positive anodes accelerate the electrons to a high velocity and focus them into a fine beam so that they all converge to a small point on the screen. Such an electrode arrangement is called an electron lens, and the development of such lenses has made the electron microscope possible. Figure 14.3 shows an

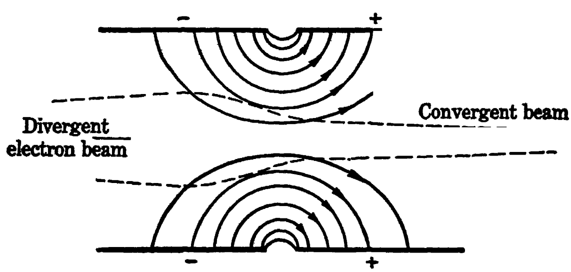


FIG. 14.3. Enlarged view of the space between the first and second anodes to show the focusing fields.

enlarged view of the space between the first and second anodes to illustrate the focusing mechanism. The curved lines represent the direction of the electric field, with an arrowhead showing the direction of the force on a negatively charged particle. Any electron taking an off-center path through the grid aperture first experiences a converging force field in which an axial component of force accelerates it forward and an inward radial component forces it inward. A little farther along, the forces are wholly axial, and still farther out the field has a diverging action. With proper electrode geometry, however, the net effect is to converge the stream because the slowly moving electrons leaving the control electrode spend a relatively long time in the convergent field and, after their forward acceleration, spend only a short time in the diverging field. This leaves the electrons with sufficient inward velocity to overcome their natural repulsion and bring them to focus at the screen.

In the electron gun of Fig. 14.2 the voltage applied to anode No. 1 controls the focusing. Ordinarily, the final anode is connected to the metal chassis so that there will be no electric field between this anode and surrounding metal structures to influence the beam as it coasts to the screen. This makes the cathode negative to the chassis by a kilovolt or more.

Deflection System. After leaving the gun, the electron beam passes between a pair of deflection plates to control the position of the screen spot. Figure 14.4 shows a side view of these plates

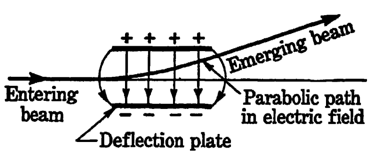


FIG. 14.4. Electrostatic deflection of the electron beam.

together with the electric field and the beam path. As soon as the electrons enter the space between the plates, they experience a lateral force and "fall" toward the positive electrode to describe a parabolic path exactly similar to the path of a horizontally fired projectile in a gravitational field. This deflects the beam and moves the resulting screen spot in direct proportion to the voltage applied between the deflection plates.

In most cathode-ray tubes two such pairs of deflection plates

operate on the beam. One vertical pair produces horizontal or X-axis deflections, and a horizontal pair provides the Y-axis vertical spot movement. This arrangement facilitates plotting the information supplied to the oscilloscope in the familiar rectangular coordinate system.

Although the effective electrostatic deflection system requires practically no power to operate it, the sensitivity is rather poor. On the average, a 1-inch spot deflection requires the application of 50 to 100 volts between the deflection plates. Consequently, the majority of input signals need amplification before they can produce a useful figure on the screen.

Since the electron beam is the equivalent of a current-carrying conductor, a magnetic field can also produce deflection. Some oscilloscopes use this magnetic deflection instead of the electrostatic type, but it is usually easier to design an amplifier to produce the required voltage across a pair of deflection plates than it is to provide the necessary current in a coil for magnetic deflection. However, magnetic deflection is always a problem with any cathode-ray tube because the stray field from nearby power-supply transformers and chokes may produce undesired deflections that show up as a ripple on the screen figure. For this reason the tube is often surrounded by a shield of high-permeability iron alloy to reduce the field strength within the cathode-ray tube.

Luminescent Screen. After focusing, acceleration, and deflection, the electron beam strikes the screen and part of the energy changes to visual radiation. This screen consists of a thin coating of finely powdered material called a phosphor applied to the inside of the tube face by spraying a volatile suspension of the phosphor or by settling it out of a liquid suspension. The thickness is carefully controlled for uniformity and to obtain a density that will produce the maximum amount of light. A thicker screen may produce more light radiation, but since the electrons strike the inside, the extra opacity may decrease the useful light reaching the front.

The screen has a twofold action. During excitation by the electrons the phosphor fluoresces and gives off light of a characteristic color, but after removal of the excitation the phosphor continues to glow—a process called phosphorescence. The glow dies off at a rate characteristic of the particular phosphor employed. This spot persistence often serves a useful purpose in

retaining the trace for a brief period of observation. Several types of screen can be obtained of which the most common are known commercially as the P1 medium-persistence green, the P2 long-persistence green, and the P5 short-persistence blue. The P1 screen is the type most often used for ordinary visual and photographic use. Most of the energy produces green light to which the eye is especially sensitive. The P2 screen has the advantage of retaining the image of a single trace for a brief visual observation, and the P5 screen is designed for good photographic efficiency and the short persistence required to prevent blurring in certain types of photographic recording.

A great many substances exhibit luminescence under electron bombardment, but the desirability of high visual efficiency and suitable color limits the choice. One of the best and most commonly used materials is zinc orthosilicate, known in the natural state as willemite, although the synthetic material produces better and more uniform results. This material gives the familiar bright-green trace of the P1 screen. Other materials provide a range of colors from red to blue; the white television screen employs a balanced mixture of phosphors to radiate the full visual spectrum.

Since the screen conducts electricity poorly, it would soon collect a negative charge sufficient to "block" or repel the beam entirely except for secondary emission. Fortunately, at moderate accelerating potentials each beam electron can produce an average of at least one secondary electron, and the screen assumes a slightly negative potential with respect to the nearest electrode. This potential stabilizes at the point where the secondary emission just equals the beam current to make the net screen current zero. To aid the collection of secondary electrons and to provide electrostatic shielding, many tubes contain a conducting coating of colloidal graphite deposited on the conical glass surface near the screen.

The image brightness depends upon the beam current, the electron-beam velocity, and the writing rate or speed with which the spot moves across the screen. Because of the mutual repulsion between the beam electrons the focus is never perfect; in fact this definitely limits the usable beam current for a given trace width. The only other way of obtaining a high-intensity image is to increase the accelerating potential. For repeating wave forms, where the beam follows the same path repeatedly, an accelerating

potential of 1 kilovolt or so will produce a sufficiently bright image. For the photography of single high-speed traces, considerably higher voltages are necessary.

14.2 The Cathode-ray Oscilloscope

The term cathode-ray oscilloscope describes an instrument incorporating a cathode-ray tube together with the necessary amplifiers and voltage supplies to make it useful. Figure 14.5 shows the block diagram for a basic oscilloscope.

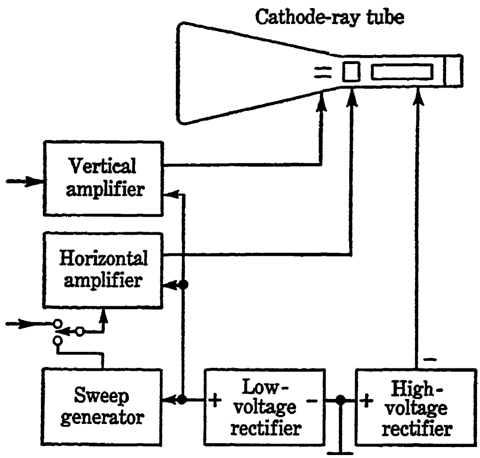


FIG. 14.5. Block diagram of a cathode-ray oscilloscope.

Vertical Amplifier. The vertical amplifier serves to amplify signals for application to the Y-axis deflection plates. It may consist of only a single stage of pentode amplification requiring a signal of about 1 volt for full beam deflection, or it may be a more complex multistage wide-band amplifier capable of amplifying very small signals to the required deflection voltage. Even the simplest oscilloscope amplifiers handle a frequency range from about 20 to 100,000 cycles; the more complex (and more expensive) ones handle a band several octaves wider.

The vertical amplifier normally contains a gain or volume control to adjust the height of the screen figure to the desired size.

Horizontal Amplifier. The horizontal amplifier provides the amplified voltage for the X-axis deflection plates. The amplifier input contains a switch used to select an input signal from either the

saw-tooth sweep generator or any external signal desired for the horizontal deflection.

Sweep Generator. The sweep generator provides a saw-tooth wave form for the purpose of deflecting the electron beam across the horizontal axis at a uniform rate. The most common circuit is essentially that of Fig. 7.16, which shows a thyratron circuit for voltage saw teeth. Usually two separate controls adjust the saw-tooth frequency—a coarse control that selects different capacitors and changes the frequency in large jumps, and a fine control rheostat (part of R in Fig. 7.16) that provides smooth control over the intermediate range between coarse steps. Thyratron sweep generators can be made to operate at frequencies from below 1 cycle per second up to about 30 kilocycles.

Low-voltage Supply. This rectifier, with negative grounded to the chassis, supplies the positive voltages needed for the amplifiers and sweep generator.

High-voltage Rectifier. This rectifier supplies the accelerating anode voltages for the electron gun and operates with the positive terminal grounded. The accelerating potential depends upon the type of cathode-ray tube. One kilovolt is a typical value for ordinary oscilloscopes, but high-intensity tubes for the photographic recording of a single sweep may require 10 kilovolts or more.

Since the electron gun draws a very small current, a simple smoothing capacitor provides sufficient filtering of the rectifier output. A tapped adjustable resistor network connected across the high-voltage rectifier output gives the different electrode voltages required by the gun.

14.3 Observation of Repeating Wave Forms

The display of a repeating wave form plotted as a function of time is one of the commonest applications of the cathode-ray oscilloscope. This type of visual presentation shows the wave form in rectangular coordinates with the variable under observation plotted against a linear time scale.

To obtain this type of display the repeating input signal is applied to the vertical amplifier. This produces a vertical spot movement on the screen in direct proportion to the instantaneous input voltage. To draw out the wave form the spot must also move horizontally across the screen at constant velocity. This

requires a horizontal deflection voltage that increases uniformly with the time, in other words, a voltage wave with a constant slope such as the saw-tooth wave form.

Ordinarily a single sweep across the screen with the spot tracing out the wave would not be particularly useful for direct viewing. The short persistence of the eye and of a normal oscilloscope screen shows the curve as a single bright flash, but any sort of detail is hard to observe. A better plan is to repeat the sweep over and over, each time plotting the curve in the same position. With a sweep frequency above 20 cycles per second the normal persistence of vision blends the images together (as in motion pictures), and the figure seems to be constantly under observation.

Figure 14.6 shows the means of doing this. Part *a* shows the

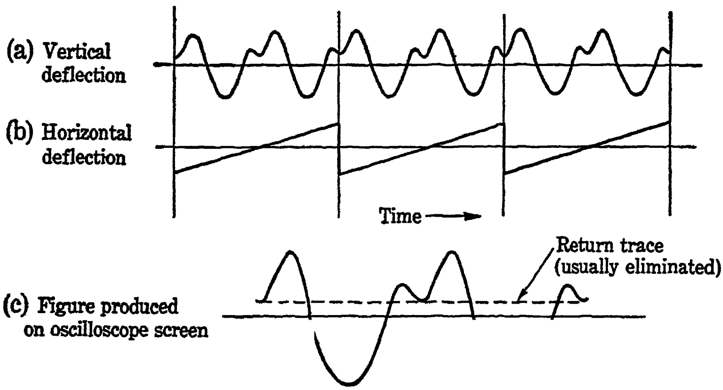


FIG. 14.6. This shows how a saw-tooth wave form applied to the horizontal deflection plates of a cathode-ray tube permits continuous observation of a repeating wave form.

wave under observation, and part *b* shows the saw-tooth wave form of *exactly half* the frequency. During one saw-tooth cycle the spot starts from the left side of the screen and traces out two complete cycles of the desired wave. Then the saw-tooth voltage collapses and returns the spot to the initial position and again traces out the identical figure. In practical circuits the saw tooth is never perfectly vertical so a tiny portion of the wave is lost, and the returning trace makes a faint line on the screen, shown dotted. The better oscilloscopes apply a momentary negative bias to the control electrode to extinguish the beam and remove the return trace.

To produce a steady figure on the screen the sweep frequency must be an exact submultiple of the input signal frequency. If it is not, the spot traces slightly more or less than an integral number of cycles and the next cycle repeats at a slightly different point on the wave form. This causes the wave to move across the screen at a rate depending on the amount that the sweep is out of synchronism. To provide this exact correspondence between the input and the sweep, the circuit feeds a small fraction of the amplified signal to the grid of the sweep generating thyratron. This causes time variations of the critical firing potential and "locks" or synchronizes the saw-tooth frequency. A synchronizing control provided on the face of the instrument adjusts the amount of input signal injected into the sweep circuit.

14.4 Curve Plotting

Many cathode-ray oscilloscope applications require the plotting of one variable in terms of another, rather than against time. In mechanical engineering, for example, pressure-displacement indicator diagrams are used to compute the developed power and to observe the operation of a reciprocating steam engine or internal-combustion engine. Mechanical arrangements work satisfactorily at slow speeds, but the cathode-ray tube is not limited by speed and, furthermore, constantly displays the indicator diagram while adjustments are made on the machine.

The figure is obtained by inserting a small pressure-sensitive plug into the cylinder head; this plug develops an electrical signal proportional to the instantaneous pressure for application to the vertical deflection amplifier. A horizontal signal proportional to the piston displacement is obtained from another electromechanical transducer connected to the crankshaft. This might consist of a resistance wire with a moving contact mechanically arranged to duplicate the piston movement. A steady current through the wire produces a voltage at the moving contact that is directly proportional to the displacement.

The two signals combined on the oscilloscope screen provide electron-beam displacements directly proportional to pressure and displacement, and the spot repeatedly travels around a path describing the desired diagram. Engine adjustments instantly show on the screen, and photographic records can be made for later analysis.

14.5 Phase Measurement

The cathode-ray oscilloscope also provides a convenient method of observing the phase angle between two sine waves of the same frequency. Figure 14.7 shows two such waves and the resulting

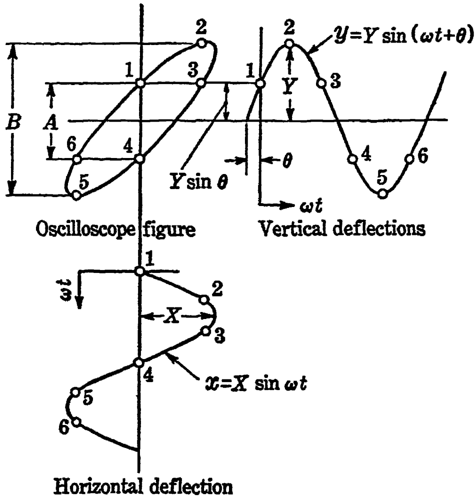


FIG. 14.7. Graphic development of a Lissajous figure showing the phase angle between two sine waves of the same frequency.

figure obtained on the screen. The small numbers represent corresponding time moments on the three figures. In the diagram the X-axis sine wave is taken as the reference with the Y-axis wave shown leading by an angle θ. At the moment of zero horizontal deflection the vertical wave has already progressed through an angle θ and reached a height of $Y \sin \theta$ (point 1). Thus the intercept of the ellipse with the vertical center axis is $Y \sin \theta$, whereas the maximum height of the figure equals Y. The ratio of these two easily measured distances gives $\sin \theta$ from which the angle can be determined. In practice, using the double intercepts A and B reduces the errors caused by any slight decentering of the figure with respect to the cross-section marks on the screen. Thus sine θ equals A/B. Figure 14.8 shows the figures obtained for a number of phase shifts between zero and 180 degrees.

This phase-measuring system suffers from two defects. (1) There is an ambiguity as to whether the angle leads or lags by angle θ.

Actually, the spot travels around the ellipse in a different direction in either case, but this is not conveniently observable. Fortunately, in many cases the decision as to lead or lag can be made

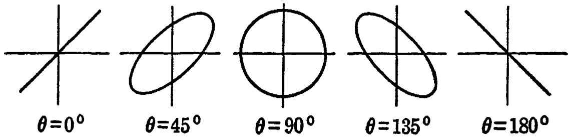


FIG. 14.8. Lissajous' figures obtained for a number of different phase angles.

from other circuit information. (2) The accuracy of the system is never particularly good, and it is especially poor for angles approaching 90 degrees where the sine changes slowly with the angle.

14.6 Frequency Comparison—Lissajous' Figures

Still another application of the cathode-ray oscilloscope is the comparison of two frequencies for the purpose of measurement or calibration. This is done by applying one frequency to the vertical and one to the horizontal deflection terminals. The resulting figure (named for Lissajous, a French mathematician who studied the curves a long time before cathode-ray tubes appeared) quickly shows the frequency ratio.

Figure 14.9 shows a number of these figures together with the corresponding frequency ratios. With unity ratio the figure ap-

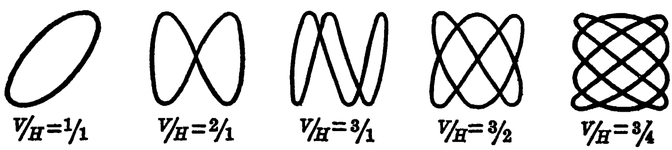


FIG. 14.9. Lissajous' figures obtained for various frequency ratios between the vertical and horizontal sinusoidal deflections.

pears as an ellipse whose fatness depends upon the phase position of the two waves. If the ratio differs slightly from unity, the two waves constantly slip in phase position and the figure slowly changes through the cycle of figures shown by Fig. 14.8.

With a frequency ratio of two, the spot makes two vertical traces for a single horizontal movement. The effect is exactly as if two complete sine-wave cycles were drawn on a transparent sheet which was then rolled into a vertical cylinder viewed from

the side. Different relative phase positions make the figure look different, but in every case the number of peaks at the top of the figure divided by the number of excursions to the side gives the frequency ratio. It is convenient to think of the figure bounded by a rectangular box to aid in counting the vertical and horizontal movements. In certain phase positions several traces may overlap to give a false count of the vertical and horizontal traces. It is best to have the figure slowly changing to avoid these particular positions.

14.7 Single-sweep Transient Recording

Many transient phenomena occur only once and can be plotted only once on the oscillograph screen. Figure 14.10, showing a

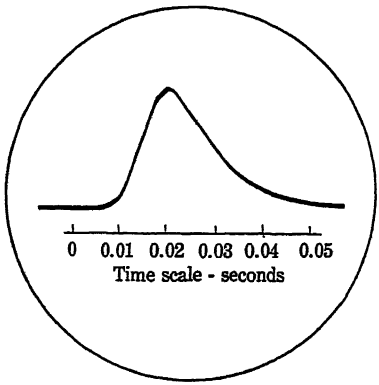


FIG. 14.10. Single-sweep recording of the relative light output of a photoflash bulb.

recording of the light output of a photographic flash bulb, illustrates an example of this type. For this record the vertical amplifier receives the electrical output of a photocell arranged to observe the light from the flash bulb. The horizontal sweep is initiated at the moment of firing the bulb, and a single rapid trace plots the curve. To obtain a photographic record, the camera views the screen through a darkened tube with the shutter opened for time exposure.

A number of cathode-ray oscilloscopes on the market (usually the more expensive kind) include facilities for producing a single sweep and adjustable delay circuits to control the time period between the start of the sweep and the beginning of the transient.

To obtain an intense trace for photographing high-speed transients requires a high-velocity electron beam produced by a tube with a high accelerating voltage. This in turn demands a bulky, high-voltage rectifier; for this reason oscilloscopes of this type are often constructed in two units, one containing the power supply and the other housing the cathode-ray tube and amplifiers. Equipment of this type also applies a negative bias to the control electrode to extinguish the beam except during the moment that the spot moves across the screen. This prevents fogging the film and burning the screen by the intense stationary spot.

14.8 Timing by *Z*-axis Modulation

The term "*Z*-axis modulation" refers to the process of turning the beam on intermittently to produce a dotted trace that is useful for timing the events represented on the curve. Figure 14.11

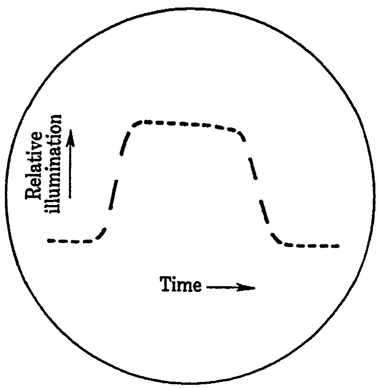


FIG. 14.11. Single-sweep recording with 500-cycle "*Z*-axis" modulation. The curve shows the light passed by a camera shutter set for an exposure of 1/50 second.

shows this type of recording used with a single sweep to record the light received by a phototube during the period of opening and closing a camera shutter set for an exposure of 1/50 second. With 500-cycle *Z*-axis modulation the distance from the beginning of one dot to the start of the next one represents 0.002 second. The record shows that the shutter stayed fully open for 0.018 second and that the total time from the start of opening to the final closing took 0.032 second. The effective exposure time, defined as the time an instantaneous shutter must stay open to pro-

duce the same exposure, can be computed by dividing the area under the curve by the maximum ordinate.

A sinusoidal voltage applied to the control electrode of the cathode-ray tube turns the beam on and off smoothly and does not produce sharply defined dots. The square wave of Fig. 14.12

FIG. 14.12. Square wave for "Z-axis" modulation.

does a better job because a voltage of this shape applied to the control electrode turns the beam either fully off or on without any intermediate hazy traces. A sufficiently square wave for this purpose can be obtained by overdriving a simple amplifier with sine waves of the desired frequency.

14.9 Continuous Recording

Sometimes a record is desired of a large number of cycles of a nonrepeating wave. This can be done by combining a cathode-ray oscilloscope with a special camera designed to focus an image of the screen on a continuously moving film strip. As shown by Fig. 14.13, the cathode-ray spot moves only in the vertical direc-

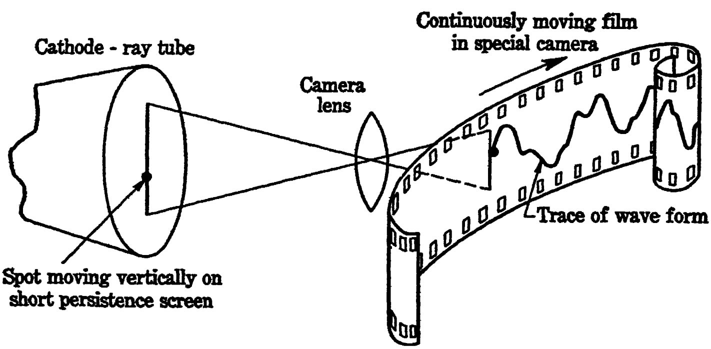


FIG. 14.13. Continuous recording with a cathode-ray oscilloscope.

tion while the film moves horizontally to provide the time axis. Thus the spot traces out the pattern showing the wave form under study. To prevent image blurring the screen must possess a very short persistence, such as that of the P5 screen. Of course the film

requires photographic processing before the image can be studied; this is a disadvantage common to all photographic recording systems.

PROBLEMS

14.1 The deflection sensitivity of a cathode-ray tube is often expressed as the effective sinusoidal deflection voltage required to produce a 1-in. peak-to-peak screen deflection. Compute the amplification required for a deflection amplifier to make each inch of screen deflection correspond to a 1-volt change in the instantaneous input potential. The tube has a deflection sensitivity of 25 rms volts per inch.

14.2 A cathode-ray tube is employed to observe the phase angle between the applied voltage and the current flowing into an impedance consisting of resistance and capacitance in series. Readings on the screen show $A = 2.3$ in. and $B = 4.0$ in. The applied voltage of 10 volts produces a current of 50 ma. Compute the circuit resistance and reactance.

14.3 Compute the effective exposure time represented by the oscillogram of Fig. 14.11.

CHAPTER 15

TRANSDUCERS

THE TERM transducer describes a device for translating a signal from one form into another. A cathode-ray oscilloscope is one type of transducer; it transforms electrical signals into a useful visual or photographic pattern. A loudspeaker transforms electrical energy into sound, while a microphone performs the reverse process. A phonograph pickup is also a familiar type of transducer; it accurately translates the needle motion into an electrical form for amplification.

Most electronic measuring and reproducing systems employ both transducers and electrical circuits. First an input transducer, or pickup, provides an electrical signal representing the physical quantity observed. This signal is then amplified, operated upon by electrical circuits, and transmitted to the desired point. Finally an output transducer, or recorder, transforms the electrical signal into the most convenient form for the particular operation. Although it is often possible to obtain a similar result without the intervening electrical system, the electrical link usually provides the advantages of flexible transmission, amplification, and speed.

The purpose of this chapter is to illustrate the possibilities of electrical measurement by discussing a few of the many types of transducers and associated equipment.

15.1 General Principles of Transducers

Transducers employ practically every electrical phenomenon known to physics. Some types have been more successful than others, but the art is still young and probably some of the neglected methods may later become important. At the risk of overgenerality, the field may be boldly divided into those which directly generate a voltage and those which depend upon the variation of some circuit parameter.

Transducers Producing a Generated Voltage. Transducers which produce a generated voltage can be subdivided into four classes: (1) magnetoelectric, (2) piezoelectric, (3) thermoelectric, and (4) electrochemical.

Magnetoelectric devices operate on the basic principle that a moving wire in a magnetic field generates an electromotive force. Since the generated voltage is directly proportional to velocity, this type of transducer is well adapted to instruments for observing the relative velocity of a mechanical system. A dynamic or moving-coil microphone employs this principle. The varying sound pressure moves a diaphragm carrying a light coil of wire in a magnetic field. Electrical amplification makes up for the very low efficiency of the device.

As explained in a following section (15.5), simple electrical circuits can perform the mathematical operations of integration and differentiation. Therefore it makes little difference whether a transducer observes displacement, velocity, or acceleration. The output of a velocity-responsive device, for example, can be made to represent the acceleration by passing it through a differentiating network.

Piezoelectric pickups take advantage of the voltage generated in a crystal under stress. Depending upon the cut and mounting of the crystal, the resulting device may be either pressure or displacement sensitive. The common crystal phonograph pickup operating on this principle employs a Rochelle salts crystal mounted as a cantilever beam. The vibrating needle slightly bends the crystal back and forth and produces an output voltage as large as 1 volt from an ordinary record. Rochelle salts has the disadvantages of solubility in water, high temperature coefficient, inability to stand high temperatures, and low mechanical strength. For these reasons the much less sensitive but more stable quartz must be used in some applications.

A variation of the phonograph pickup is the surface analyzer. The minute vibrations of a jewel stylus rubbed across a surface produce an electrical output which is amplified to operate a meter calibrated to read the surface roughness.

The *thermoelectric* effect provides a convenient means of producing an electrical output proportional to the temperature. The familiar thermocouple consisting of the junction of two dissimilar metals is one of the most accurate and flexible methods of measur-

ing temperature. Since a thermocouple can p
current, the tiny output voltage is usually indic
meter or read by a sensitive potentiometer.

Electrochemical processes also play a part in the sc
urement. The glass electrode coupled with the vacuum
meter provides a quick and convenient method of meas
hydrogen-ion content of solutions. The electrode produces
age proportional to the ion concentration, but its resistance
high (above 10^8 ohms) that even a vacuum-tube circuit must
carefully designed to reduce the current drawn from the electrode.
This is quite different from the thermocouple, which can directly provide the power for operating a meter.

Transducers Varying a Circuit Parameter. Many transducers do not inherently generate an output voltage but vary instead some electrical circuit parameter such as resistance, inductance, or capacitance. This variation can be observed by a null method such as a potentiometer or a bridge, or it may be used to produce a varying voltage suitable for amplification.

Resistance variation may be applied to either temperature measurement or the observation of mechanical displacement. The resistance thermometer takes advantage of the temperature coefficient of resistance of metals; nickel and platinum make stable resistance elements for this purpose. Such a temperature-sensitive resistance element coupled to an automatic recording bridge provides an excellent means of obtaining a continuous temperature record. Resistance elements for the measurement of strain have a shorter history. The most common ones depend upon stretching a fine resistance wire. This increases the length, reduces the cross section, and affects the electrical properties of the material. As a result, the resistance changes in proportion to the strain; this can be observed with a Wheatstone bridge.

Inductance elements also serve a useful part in electrical measuring instruments. One example is the gage for measuring the thickness of a coating applied to a flat steel surface. The instrument, as shown in Fig. 15.1, consists of an iron yoke wound with a coil of wire. Placing it on a steel surface completes the magnetic circuit except for two gaps where the flux passes through the coating material. A thick gap increases the reluctance of the magnetic circuit and lowers the coil inductance. With a suitable electrical circuit a calibrated meter can read the gap thickness directly.

The *condenser* microphone is another example of circuit-param-

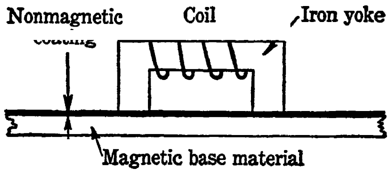


FIG. 15.1. An inductance-variation type of transducer which measures the thickness of a nonmagnetic coating on a magnetic material. This arrangement is used commercially for measuring enamel thickness on sheet steel.

eter variation. In this device a stretched metal diaphragm moves in response to the minute pressure variations caused by the sound waves. This movement produces variations in the electrical capacitance between the diaphragm and a rigid back plate parallel to and slightly spaced from it. Applications of this principle include the dynamic measurement of pressure in internal-combustion engines and guns and the recording of explosive blast pressures.

15.2 The Resistance Strain Gage

One of the transducers important enough to discuss in more detail is the resistance strain gage. Although the principle is simple, the gage went through a development period of about 10 years in the 1930's; since 1940 its use has increased, until it is now a standard item in any mechanical research program.

As illustrated by Fig. 15.2, the gage consists of a length of fine

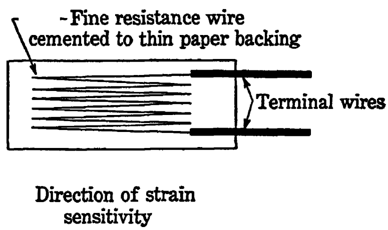


FIG. 15.2. The resistance strain gage—bonded type.

wire about 0.001 inch in diameter bent back and forth and cemented to a thin tissue paper. The paper serves to keep the wire strands in place, to support the heavier connecting leads, and to insulate the wires from the surface to which it is attached. In use the gage is cemented directly to the surface of the structural member under test so that it experiences the same strains as the surface. The light flexible assembly has a negligible effect on the motion of any reasonably heavy structural member.

Suitable adhesives include the ordinary household celluloid cements and a number of similar cements designed particularly for this purpose; they provide good adhesion and maintain the electrical insulation. Surprisingly enough, the bond is so good that the gage responds to compression as well as to tension. This results from the use of extremely fine wire to provide a large surface area compared with the cross section. Thus, although the wire stress is high, the adhesive force required per unit of surface area is low.

The gage operates on the principle that stretching a wire changes its resistance. The resistance change occurs because (1) the wire stretches, (2) the diameter shrinks, and (3) the specific resistivity of the wire material changes. These factors affect the resistance of a wire according to the relation

$$R = \frac{\rho l}{A} \tag{15.1}$$

Since an increase in length produces a decrease in area, the fractional resistance change exceeds the strain experienced by the wire. Fortunately, many metals and alloys also experience an increase in resistivity ρ under tension, which also contributes to increasing the sensitivity of the gage. This ratio between the resistance change and the strain is called the gage factor, defined as

$$\text{Gage factor} = \frac{\Delta R/R}{\Delta l/l} \tag{15.2}$$

In practical gages this factor usually exceeds 2. Before the availability of commercial gages each individual winding had to be calibrated to determine the resistance and factor, but commercial elements can now be obtained with predetermined characteristics held to close tolerances.

The gages can be obtained in a number of different resistance values; 120 ohms is a commonly used value. They are also wound of wire with a low thermal coefficient of resistivity to minimize the effect of temperature upon the gage resistance. Any residual temperature changes can be balanced out with an unstressed dummy gage mounted nearby.

15.3 The Strain-gage Bridge

Since the normal working strain in a steel member is less than 0.001 inch per inch, the expected gage resistance change seldom exceeds 0.2 percent and an accurate method must be used to determine the resistance. The best arrangement for this purpose is

the ordinary Wheatstone-bridge circuit shown by Fig. 15.3. In this circuit R_x represents the strain gage, G the galvanometer for detecting the null point, and E the applied voltage. At balance

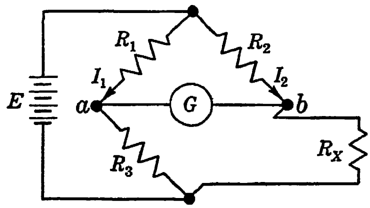


FIG. 15.3. The basic strain-gage bridge.

the sensitive galvanometer reads zero, indicating no voltage difference between points a and b. Therefore the voltage drops across R_1 and R_3 are equal.

$$I_1R_1 = I_2R_2 \tag{15.3}$$

Since the galvanometer current at balance is zero, current I_1 continues on down through R_3, and I_3 through R_x. Equating the voltage drops through these two resistors, we obtain

$$I_1R_3 = I_2R_x \tag{15.4}$$

Dividing Eq. (15.3) by Eq. (15.4), we obtain the bridge equation at balance

$$\frac{R_1}{R_3} = \frac{R_2}{R_x}$$

which solved for R_x gives

$$R_x = R_3\left(\frac{R_2}{R_1}\right) \tag{15.5}$$

A particularly simple arrangement is to make resistors R_1 and R_2 equal; in this case the changes in R_3 measure the change in R_x directly. Furthermore, it is possible to measure accurately a change of 0.1 percent in R_x with bridge resistors accurate to only 1 percent. Figure 15.4 illustrates the slight circuit change for

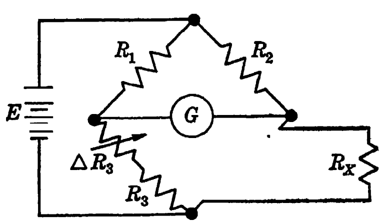


FIG. 15.4. A refinement of the strain-gage bridge to measure accurately small changes in R_x.

obtaining this result. With this arrangement the small resistor ΔR_3 is first set to zero and the bridge carefully balanced by adjusting R_3 itself. Under these conditions

$$R_x = R_3\left(\frac{R_2}{R_1}\right) \tag{15.6}$$

Now let us suppose that a strain applied to the specimen increases the gage resistance by a value ΔR_x and that balance has been re-established by adjusting the resistor ΔR_3. With this new balance

$$R_x + \Delta R_x = (R_3 + \Delta R_3)\left(\frac{R_2}{R_1}\right) \tag{15.7}$$

By multiplying out the right side of Eq. (15.7) and subtracting Eq. (15.6), we obtain

$$\Delta R_x = \Delta R_3\left(\frac{R_2}{R_1}\right) \tag{15.8}$$

Thus although the incremental changes are very small, they can be measured to the same order of accuracy that we know the adjustable resistor ΔR_3 and resistors R_1 and R_2. Since commercially available adjustable decade resistors are commonly constructed to an accuracy of 0.1 percent, the bridge can measure incremental resistance change with excellent accuracy.

The problem of temperature correction can be met by using two gages instead of one, as illustrated by Fig. 15.5. Here resistor R_2

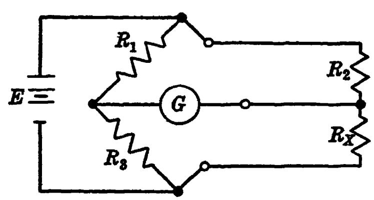


FIG. 15.5. A strain-gage bridge with a dummy gage R_2 in one arm to provide temperature compensation.

is replaced with an unstressed dummy mounted close to R_x to maintain the same temperature. As an aid, many gages come with a flannel covering cemented to the outer surface to insulate the gage wires from the surrounding air and make the gage temperature approach that of the underlying surface. With an equal-arm bridge identical changes in R_2 and R_x do not affect the balance. This can be seen by rewriting Eq. (15.5) in the form

$$R_x = R_2 \qquad (15.5)$$

Since the ratio R_3/R_1 is unity, symmetrical changes in R_2 and R_x do not affect the equality.

This arrangement also cancels out the effect of the gage connecting leads, which add to the circuit resistance and possess a relatively high temperature coefficient. This is helpful because the leads cannot always be made short; usually a number of gages at various points on a structure are brought to a common measuring instrument.

Figure 15.6 shows a clever way of making the dummy gage R_2

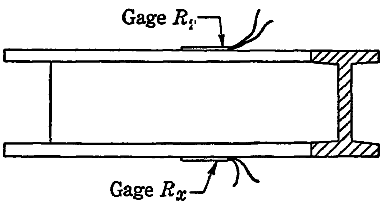


FIG. 15.6. On a symmetrical structure, gage R_2 can be placed to undergo an equal and opposite strain to that of R_x. This doubles the bridge sensitivity.

perform a useful function. With a symmetrical beam the second gage can be placed on the opposite side so that one element suffers tension while the other experiences compression when the beam is bent. This increases the resistance of one element while decreasing the other and doubles the bridge unbalance. The temperature and lead corrections remain unchanged.

Dynamic Strain Measurement. The null method of measuring the strain works nicely under static conditions, but when the deflection varies dynamically it is impossible to follow anything except very slow variations. For a dynamic strain record, exactly the same bridge can be used with the terminals a and b connected to an amplifier (Fig. 15.7). As the gage undergoes strain alterna-

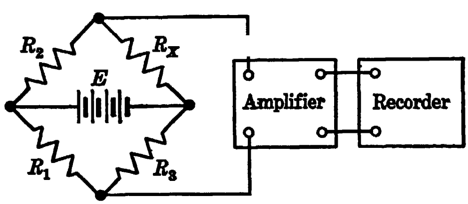


FIG. 15.7. Strain-gage bridge with an amplifier to record transient strains. It is usually more practical to use an alternating voltage in place of E.

tions the voltage observed by the amplifier varies in direct proportion and swings from plus to minus when the strain changes from tension to compression.

The alternating signal developed is only in the order of 10 millivolts, but this is easily amplified for presentation on a cathode-ray oscilloscope or a recording galvanometer. For low-frequency vibrations having a frequency of less than 1 cycle per second, however, the problem becomes more difficult because it is hard to construct R–C amplifiers with a response extending to such low frequencies. Theoretically a direct-coupled amplifier extends the response to zero frequency, but practical d-c amplifiers always exhibit instability; the possibility of slow drift causes some uncertainty regarding the position of the zero line on a recording.

Exciting the bridge with an alternating voltage avoids this predicament by providing a varying alternating voltage for the amplifier. The amplified output then appears as a modulated wave, the envelope representing the strain variations. Of course the applied frequency must be considerably higher than the vibration frequency in order to show the envelope details properly. This works nicely except for the fact that an alternating voltage does not conveniently indicate polarity and hence produces an ambiguity

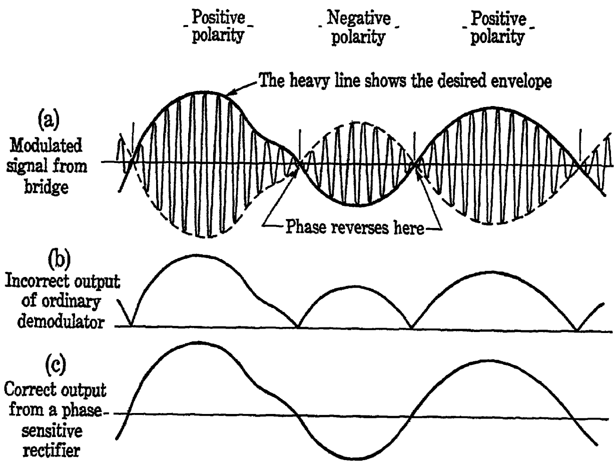


FIG. 15.8. The modulated output voltage of a strain-gage bridge operating with an a-c input. With the bridge initially balanced for zero strain, the output changes phase each time the strain changes polarity.

as to the sign of the strain. Actually, the alternating bridge output does reverse in phase as it goes through the zero point, but this is not evident to the eye viewing the oscilloscope image, nor does it affect the ordinary envelope demodulator that might be used to rectify the output to recover the envelope. Figure 15.8 explains this in more detail. Part *a* shows the modulated signal from the bridge with one envelope drawn as a heavy line to indicate the phase reversal at the point where the signal goes through zero. Part *b* shows the rectified curve obtained from an ordinary demodulator whose output depends only upon the signal amplitude. The curve contains all the information regarding the amplitude of the original strain but fails to indicate the sign. Curve *c* shows the output of a phase-sensitive circuit which produces a positive or a negative output depending on the input phase.

The phase-sensitive rectifier of Fig. 15.9 compares the signal

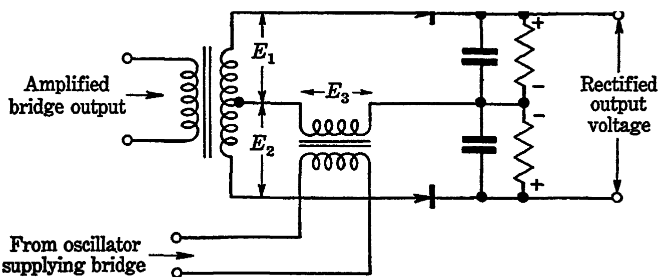


FIG. 15.9. A phase-sensitive rectifier which produces an output proportional to the input and of corresponding polarity.

to the original alternating voltage applied to the bridge by means of a pair of identical half-wave rectifiers. With no input signal from the amplifier, voltages E_1 and E_2 supplied by the center-tapped transformer are also zero and each rectifier circuit operates with the same alternating input voltage E_3. Therefore each one produces the same magnitude of d-c output voltage, and the net output voltage is zero. An input signal changes matters. Equal voltages E_1 and E_2 now appear to upset the balance. With one phase of input voltage, E_1 adds to E_3 and increases the output of the upper circuit. At the same time E_2 subtracts from E_3 to decrease the lower circuit output. With this unbalance the two d-c outputs no longer cancel, and the upper output terminal be-

comes plus. Reversing the input-voltage phase makes E_1 *subtract* from E_3, E_2 *add* to E_3, the lower output exceed the upper, and the output voltage reverse. In this way the circuit provides a rectified output proportional to the input signal and of the proper polarity.

The use of an alternating supply voltage for the bridge does complicate the balance problem because the circuit capacitances now become important, especially when long leads are used to reach to the gage location. The effect of these capacitances can be balanced out by placing an equal capacitance across the corresponding resistance in the opposite leg of the bridge. Another arrangement is to make the bridge out of four gages, all of them mounted near one another. The leads then represent the oscillator and amplifier connections, which may have to be shielded from one another, but the capacitance involved does not affect the balance. Where the arrangement of Fig. 15.6 can be employed, an especially sensitive bridge can be made with two tension and two compression arms. The two tension gages are placed in diametrically opposite bridge arms with the two compression gages in the other two positions. This provides just four times the sensitivity of a single-gage bridge.

15.4 The Differential Transformer

Another useful pickup for the measurement of mechanical movement is the differential transformer. Figure 15.10 shows this to

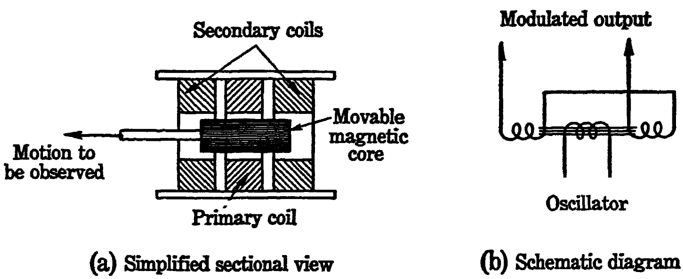


FIG. 15.10. Simplified cross-sectional view and circuit diagram of a differential transformer.

consist of three coils enclosing a movable iron core, the whole mounted in an enclosure an inch or two long. This device depends upon the mutual inductance between the energized central primary winding and the two identical secondaries.

With the movable core at dead center the system is perfectly symmetrical, and the alternating primary current induces identical voltages into the two secondaries. These cancel to produce zero output with the coils connected in phase opposition, as shown. Displacing the core to the left, however, disturbs the balance, more flux links the left-hand coil, and the output rises. A displacement to the right produces the opposite effect. Again the output rises but with the opposite phase because the voltage of the right-hand secondary now exceeds that of the left. With proper mechanical design this relation between output voltage and displacement can be made linear for a reasonable distance.

The characteristic curves of Fig. 15.11 illustrate the linearity

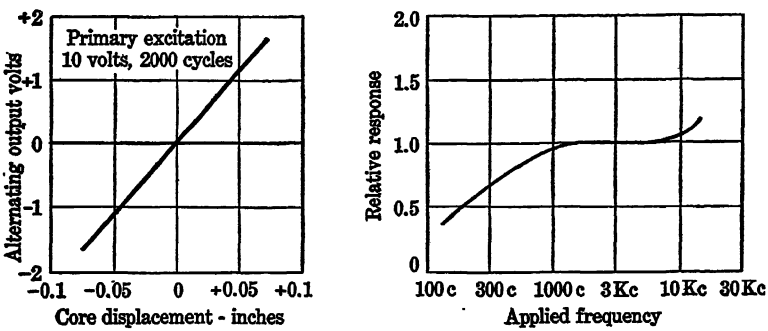


Fig. 15.11. Characteristic curves illustrating the behavior of a differential transformer.

and frequency response of a typical unit. The linearity curve indicates that the differential transformer is adapted to the measurement of larger displacements than is the resistance strain gage and that the alternating output voltage is relatively high. Commercially available units cover a considerable range of physical size and corresponding linear-displacement range. The frequency-response curve shows the wide range of frequencies suitable for exciting the primary. The use of frequencies on the flat portion of the curve has the advantage of making the sensitivity independent of the oscillator-frequency drift. However, lower frequencies are perfectly suitable as long as the exciting frequency remains well above the highest frequency component of the core motion.

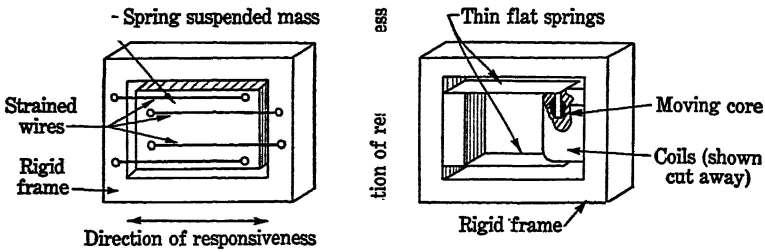
When the differential transformer is used to record vibrations,

the output appears as a modulated wave with phase reversals exactly like Fig. 15.8. Amplification followed by demodulation with the phase-sensitive rectifier of Fig. 15.9 provides an electrical signal that faithfully follows the original motional variations.

15.5 The Accelerometer

In the field of vibration study, the acceleration of a structure is one of the most important variables to observe. To measure displacement directly requires some fixed reference point; this is obviously impossible when testing the flutter of an airplane wing during flight. The acceleration of an object, however, can always be observed by the accelerating force required, without the need for any additional reference information.

Figure 15.12 illustrates two types of accelerometers, one em-



(a) Strain-gage type (b) Differential transformer type

FIG. 15.12. Simplified sketches of two types of accelerometers for use in vibrations studies.

ploying the strain-gage principle and the other based on the differential transformer. The basic element of each consists of a mass mounted on a spring suspension that permits motion in one direction only. In the strain-gage type four fine strain wires provide the majority of the spring tension, although an additional spring support prevents motion in all directions not parallel to the wires. The wires are prestressed to remain always in tension. The force required to accelerate the suspended mass changes the wire strain (two increase and two decrease), which causes unbalance of a strain-gage bridge. Actually, the accelerometer contains the four identical elements for a complete bridge and is thus automatically compensated for temperature changes.

The second accelerometer also contains a spring-suspended mass, in this case the movable core of a differential transformer. Here

too, acceleration in the correct direction (or the component of acceleration in that direction) causes a relative movement between the core and the rigid frame and thus affects the electrical output.

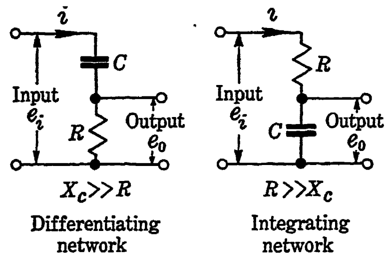
Unfortunately the spring-mounted mass is a mechanically resonant system, and vibrations in the vicinity of the resonant frequency cause deflections all out of proportion to the applied motion. This can be corrected by filling the case with a viscous liquid to damp the motion. The new inert silicone oils possess a very low temperature coefficient of viscosity, which makes them especially well suited to this purpose. Analysis shows that a little less than critical damping produces the best results. With this damping the response to sinusoidal acceleration is uniform up to about half the resonant frequency. Above resonance the suspended mass tends to stand still while the rest of the instrument vibrates about it. Above twice the resonant frequency the unit measures displacement directly.

Available commercial accelerometers possess resonant frequencies from perhaps 10 to 1,000 cycles per second. The stiffer high resonant-frequency units are naturally less sensitive than those designed for a low natural frequency. This is not a great disadvantage because high-frequency vibrations demanding an accelerometer with a high limiting frequency usually involve greater accelerations than do slow low-frequency movements.

Electrical Integration and Differentiation. It is possible to construct simple electric networks to perform the mathematical operations of integration and differentiation. This makes it simple to take the output of an accelerometer, for instance, and by integration convert it to a signal representing the velocity. Conversely, differentiating the output of a position sensitive transducer gives the same signal that would have been obtained from a velocity-sensitive type.

Figure 15.13 shows an integrating and a differentiating network together with a brief analysis of each circuit. The circuits are not perfect; each involves the approximation that the majority of the applied voltage appears across one of the two circuit elements. In the differentiating circuit, for example, the voltage drop across R must be small so that the applied voltage closely equals the voltage across C. This means that the circuit always involves a large voltage sacrifice that can be made up by amplification. It also limits the useful frequency range of the circuit, be-

cause a sufficient high frequency can always be found at which the capacitive reactance approaches the value of the resistor. Thus the differentiating circuit will cover from zero frequency up



$$i \cong \frac{e_i}{\ }$$

$$i = C\frac{de_i}{dt} \qquad i = C\frac{de_0}{dt}$$

$$e_0 = iR \qquad C\frac{de_0}{dt} = \frac{e_i}{R}$$

$$\therefore e_0 = RC\frac{de_i}{dt} \qquad de_0 = \frac{e_i}{RC}\,dt$$

$${}_0 = \frac{1}{RC}\int e_i\,dt$$

FIG. 15.13. Simple electrical differentiating and integrating circuits together with an analysis of their behavior.

to a limiting value determined by the circuit design. The integrating circuit, on the other hand, can cover all frequencies *above* that for which R is several times the value of X_c.

15.6 Recording Elements

The end product of the systems we have been discussing is usually some type of recording that displays the information as a curve plotted against time. The cathode-ray oscilloscope serves admirably for an immediate visual interpretation, and with photographic accessories the image can be recorded for further study. In fact, for high-frequency phenomena it is the only device with sufficient speed to record the wave form. For the lower frequencies extending up to perhaps 10 kilocycles, however, there are electromechanical oscillographs that possess advantages of directness, accuracy, and portability that make them superior to cathode-ray equipment.

The mechanical oscillographs that record on photographic film or paper consist essentially of a refinement of the ordinary moving-coil d-c meter to permit rapid response to current changes. Fig-

ure 15.14 shows two such types of recording galvanometers. In Fig. 15.14*a* the moving coil is reduced to a pair of tightly stretched

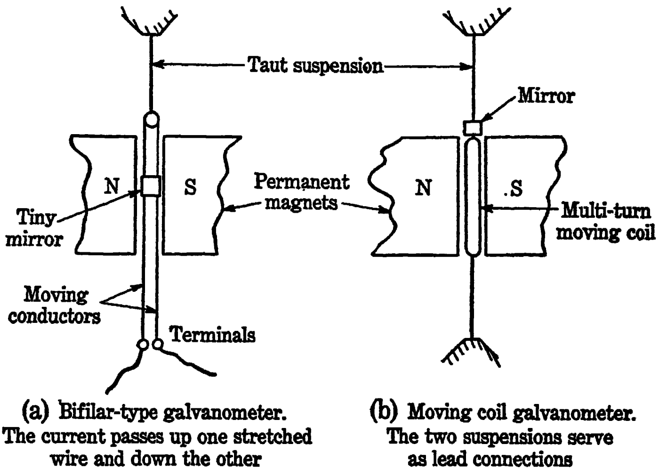


FIG. 15.14. The two basic types of galvanometer elements used in mechanical recording oscillographs.

wires placed between the poles of a strong permanent magnet. Cemented to the wires a tiny mirror rotates slightly about a vertical axis when the current passes up one wire and returns through the other. The low mass and high tension make the resonant frequency high and permit the moving system to respond rapidly. Damping to eliminate the natural vibrations of the element is obtained by immersing the whole moving element in a transparent viscous oil. Elements of this type can be designed with a flat frequency response up to 10 kilocycles per second, but their sensitivity is rather low. Reducing the tension increases the sensitivity and simultaneously reduces the useful frequency range. In fact this type of element never has particularly good sensitivity because the moving element is only the equivalent of a single-turn coil. A typical sensitivity for one of these galvanometers with a resonant point of 3,000 cycles per second is 50 milliamperes per inch deflection obtained on the film. This sensitivity is approximately inversely proportional to the resonant frequency squared; for example, a galvanometer with a 1,000-cycle resonant frequency would produce a 1-inch deflection with about 5 milliamperes.

The more sensitive galvanometers employ a multiturn coil (Fig.

15.14b) which, of course, increases the mass and reduces the maximum usable frequency. Depending upon the number of turns and the stiffness of the suspension, this type of construction results in resonant frequencies from as low as 30 up to about 1,000 cycles per second. A 30-cycle unit may produce a 1-inch deflection with a current of only 5 microamperes flowing through the coil.

Although the galvanometer is the heart of the recording system, the complete apparatus must include a light source, an optical system, and some sort of drum or magazine to hold the photographic film or paper. Figure 15.15 shows a simplified layout of

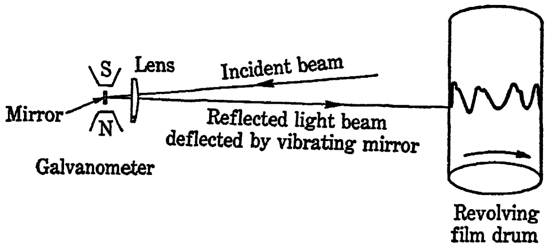


FIG. 15.15. The basic elements of a recording oscillograph.

the recording system. The photographic record is made by a narrow light beam falling upon the moving photographic emulsion. Mirror oscillations deflect the light beam in proportion to the galvanometer current, and the continuous motion of the photographic material provides the time scale against which the resulting curve is plotted.

A complete oscillograph usually includes a number of galvanometers to plot simultaneous traces on the same record and some sort of timing system to mark a time scale on the chart. Figure 15.16 shows a record taken with such an instrument using only a single galvanometer to record the output of an accelerometer. The oscillogram records a drop test with the accelerometer fastened to a mass that is suddenly dropped a short distance into a cushion. At the moment of release the acceleration suddenly rises from zero to that of gravity and remains substantially constant during the free fall. The initial overshoot shows a slight underdamping of the moving element. At the bottom of the fall the rapid deceleration momentarily throws the trace off the record, but the return of the trace to a positive value of about gravity indicates that the mass bounced at least once before coming to rest.

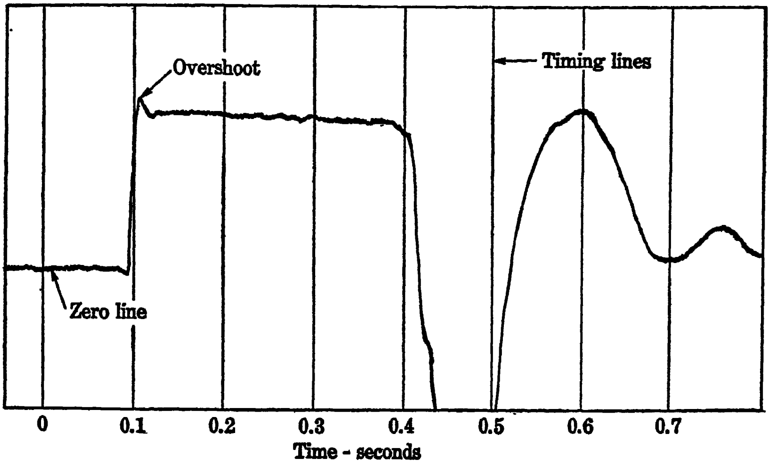


FIG. 15.16. An oscillographic record taken of an accelerometer drop test.

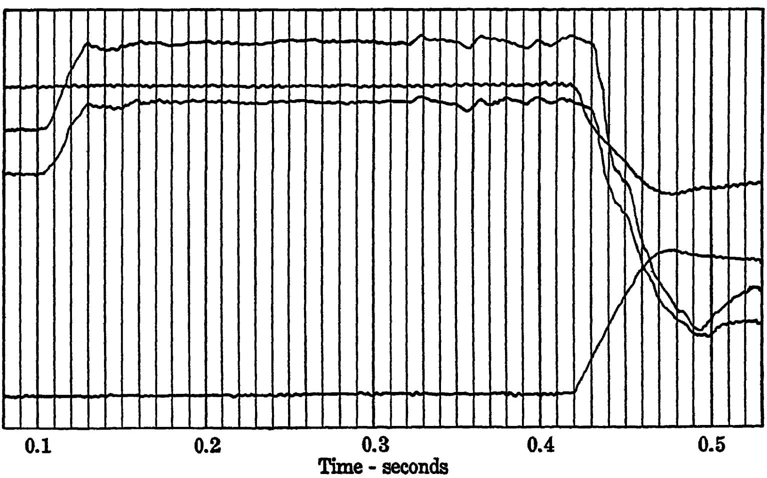


FIG. 15.17. Multielement oscillographic record of a drop test on an aircraft landing gear.

A flashing light source records the vertical timing lines simultaneously with the oscillogram. On Fig. 15.16 each line represents 0.1 second, while on Fig. 15.17 each heavy line represents 0.1 second with lighter 0.01-second intervals marked in between. The other traces on Fig. 15.17 show accelerometer and displacement

records taken during a drop test on an aircraft landing gear. Another timing method is to record a sinusoidal trace from an accurate 100-cycle oscillator or from the 60-cycle line.

The chief disadvantages of the multichannel recording oscillograph described above are the delay for photographic development of the record and the fact that a good oscillograph costs several thousand dollars. This cost factor is one reason for the use of cathode-ray equipment for applications where it is satisfactory.

To meet the need for a direct recording system, there have appeared on the market several types of oscillographs specifically designed to produce an instantaneous permanent record on a moving paper chart. Some of these actually operate a rapidly moving pen that traces an ink record on the paper. Others use an electrically sensitive paper that darkens when a current-carrying stylus passes over the surface. Such instruments are limited in speed of operation because of the relatively massive pen or stylus that must move to trace out the record. It is perhaps remarkable that the better instruments of this type can accurately record frequencies from zero up to about 100 cycles per second.

PROBLEMS

15.1 A 120-ohm resistance strain gage with a gage factor of 2 is used in an equal-arm bridge initially balanced before the application of strain. Rebalance after straining the specimen requires a change of 0.1 ohm in R_3. Compute the strain.

15.2 A strain-gage bridge employs two equal gages in adjacent bridge arms arranged to experience equal and opposite strains. The gages have factors of 1.8, the applied d-c potential is 24 volts, and the two bridge output terminals connect to an amplifier, as in Fig. 15.7. After initial balancing at zero strain, an applied strain of 0.001 produces a small output voltage. Compute the value of this voltage.

15.3 A differential transformer with the characteristics of Fig. 15.11 operates with a primary excitation of 15 volts at 2,000 cycles. The core has an initial displacement of 0.01 in. from center plus an alternating displacement having a peak-to-peak value of 0.005 in. at a frequency of 10 cps. (*a*) Sketch the output voltage wave. (*b*) Determine the percentage modulation. (*c*) Compute the effective value of the carrier amplitude.

15.4 Compute the distance of free fall experienced by the accelerometer giving the record of Fig. 15.16.

CHAPTER 16

THE VACUUM–TUBE VOLTMETER

THE FUNCTION of a vacuum-tube voltmeter is to measure the voltage difference between two points in a circuit without drawing an appreciable current from it. Ordinary moving-coil instruments do require current for operation, and there are many times when drawing even 1 microampere may seriously affect the observed voltage. This is especially true in chemical research where the potential difference between a pair of electrodes is used to measure ion concentration or to observe the end point in a reaction. In most cases, drawing even the slightest current from the electrode system may seriously affect the accuracy of the results.

A vacuum-tube voltmeter takes advantage of the grid-control possibilities of a tube—control obtained without an appreciable current flow to the control electrode. Thus the role of the vacuum-tube is to translate, and perhaps to amplify, the voltage under test into a form suitable for operating an indicating meter.

16.1 A Basic D–C Vacuum-tube Voltmeter

Figure 16.1 shows the circuit diagram for a d-c vacuum-tube voltmeter. It is only one of many such circuits, but this one

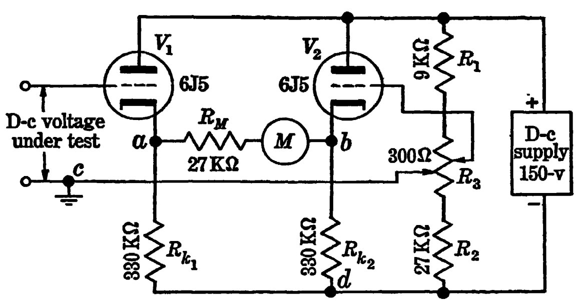


FIG. 16.1. A stable d-c vacuum-tube voltmeter based on the cathode-follower circuit.

works very well; it has the advantage of simplicity, and it is relatively easy to analyze. The term d-c voltmeter refers not to the power supply but to the ability of the circuit to measure direct voltages rather than alternating potentials. Depending upon the circuit design, this arrangement can measure voltages up to 10 volts or more without drawing an appreciable current. The circuit values shown are suitable for a meter covering a range of 0 to 3 volts. This range can be extended, at the expense of input current, by adding a high-resistance voltage divider, as discussed later.

The circuit consists of two identical cathode followers with the meter connected between the cathodes and with an arrangement for obtaining the proper bias for each tube. Ideally, with identical tubes and cathode resistors, only one bias tap would be necessary on R_3. Then with zero input voltage the two grids would be at the same potential, because of symmetry point a would have the same potential as point b, and meter M would read zero. Actually, of course, the two circuit halves cannot be exactly the same, and a slight bias adjustment on either tube must bring the meter to zero. Thus the function of resistors R_1, R_2, and R_3 is to divide the d-c supply voltage into two fixed parts and provide a slight individual bias adjustment.

As discussed in Chap. 10, a cathode follower acts to transfer a grid-voltage change over to the cathode circuit with nearly the same amplitude despite the presence of the cathode-load resistor. Thus a small voltage applied to the input produces a similar change at a. No change occurs in the other cathode follower, and point b remains at about the same potential. This causes current to flow through the meter and produce a reading proportional to the input signal.

Figure 16.2 illustrates this more accurately in terms of the simplified equivalent circuit developed from Fig. 10.8 for a single cathode

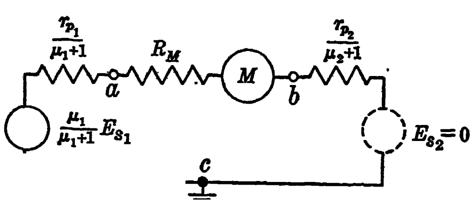


FIG. 16.2. Equivalent circuit for the vacuum-tube voltmeter of Fig. 16.1.

follower. Resistors R_{k1} and R_{k2} have been omitted from the diagram because they are large compared with the other circuit elements and have little effect. The equivalent circuit takes advantage of the fact that points c and d operate at a fixed potential difference. Therefore, as far as input voltage *changes* are concerned the two points are the same. Figure 16.2 shows the voltmeter to act as a simple series circuit with an impressed voltage slightly less than the input signal E_s. Furthermore, the fraction $\mu/(\mu + 1)$ is insensitive to changes in amplification factor, and most of the circuit resistance appears in fixed resistor R_m because the equivalent resistors $r_p/(\mu + 1)$ are relatively small. Thus tube replacements and gradual changes with time do not affect the meter calibration.

The circuit is also exceptionally linear. An input voltage that increases the plate current of V_1, for example, decreases the plate current of V_2 a nearly equal amount. This reduces the value of r_{p1} but increases r_{p2}. As a result the total series circuit resistance remains essentially constant despite the presence of nonlinear vacuum tubes.

It is also important that the circuit be unresponsive to variations in the d-c supply voltage which is usually supplied from a rectifier that follows any a-c line voltage changes. The circuit symmetry assures this independence. An increase in d-c supply voltage affects the grid and plate voltages of both tubes equally and causes equal voltage rises at a and b. This does not affect the current through meter M. Of course the two circuit halves are never exactly identical, but analysis shows that even with tubes differing by 10 percent in amplification factor any supply voltage change is reduced by a factor of about 500 in its effect on the voltage between a and b. Stabilizing the supply voltage with a glow tube and carefully picking the triodes for balance make the circuit sufficiently stable for precise measurement.

Reduction of Grid Current. The vacuum-tube voltmeter circuit of Fig. 16.1 draws a negligible grid current for most measurements, but it is sometimes necessary to take special precautions to reduce the input current to a minimum. Commercial receiving type tubes are not designed for particularly low grid current, and there are special circuits and tubes available to meet the most exacting requirements, but we shall discuss the reduction of grid current with ordinary tubes and circuits.

The most important sources of grid current are (1) leakage through the tube insulation, (2) electron flow to the grid, and (3) positive ion flow to the grid. Ordinary leakage is worst in tubes with the grid lead through the base. The closely spaced base pins, the relatively poor insulation, and the opportunity for collection of dirt and moisture makes the amount of leakage unreliable at best. The same is true of the tube socket. The selection of a glass tube with the grid lead through the top solves this problem if the glass envelope is kept clean and free of moisture. Internal leakage through the mica supports is usually small, although poor tubes occasionally occur.

Electrons always flow to a positive grid, but some manage to reach even a negative grid because the electrons leave the cathode with finite velocities sufficient to carry them against a small negative potential. The upper curve of Fig. 16.3 illustrates this and

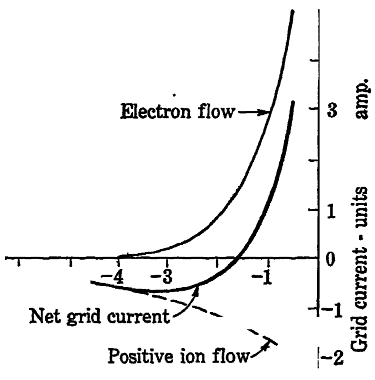


FIG. 16.3. Curves showing the two major components of grid current in a vacuum tube operating with a negative grid.

shows that even for a negative bias of 1 volt the electron current may be appreciable. The current scale shows the general order of magnitude, but this varies considerably from tube to tube. The diagram also shows a curve for the positive-ion current due to the presence of gas, the amount of which depends upon the excellence of evacuation and degassing. Collisions of plate-current electrons with residual gas produce these ions which naturally flow to the negative grid. The ion flow shown by the curve drops off, not because of less attraction by the grid, but because the negative

grid decreases the plate current and the rate of ion production. This suggests operating the tube with the lowest possible plate current and voltage to reduce the positive ion contribution to grid current. In fact, reducing the plate voltage below the ionizing potential will practically eliminate the positive ions.

At some particular bias the net grid current becomes zero, and it is here that we wish to operate the vacuum tube. The point at which the curve crosses zero is called the free-grid potential because a disconnected or "free" grid immediately seeks this potential. Should it initially be slightly more negative, for example, the excess positive-ion flow rapidly increases its potential to the free-grid point.

We can easily adjust the vacuum-tube voltmeter of Fig. 16.1 for operation at this point by temporarily disconnecting the grid lead to V_1 right at the tube. (An appreciable length of wire left hanging on the grid terminal may receive induced voltages from stray alternating fields.) The grid will immediately assume the free-grid potential causing meter M to read. Then M is set to zero by readjusting the grid-bias control of V_2. The next step is to reconnect the grid, short the input terminals, and adjust the bias control of V_1 to restore the zero reading. Since the bias for V_2 was adjusted to zero meter reading with V_1 at the free grid potential, the grid of V_1 must again be operating at the zero current point although it is now electrically tied to the remainder of the circuit. If now a small input voltage is measured, the grid will operate in the vicinity of the minimum current point and have the least possible effect on the voltage under test. Actually, of course, applying an input voltage does shift the grid operating point a little, but since the cathode voltage follows closely behind, the grid-to-cathode voltage changes little.

Extension of Range. For many applications the precautions just discussed regarding the reduction of input current are unnecessary, and any voltmeter having a resistance higher than several megohms will serve. Under these circumstances the instrument range can be extended by the use of a high-resistance voltage divider, illustrated by Fig. 16.4. Taps at 0.1, 0.01, and 0.001 part of the total resistance extend the range by factors of 10, 100, and 1,000. For a well-designed circuit the total divider resistance can be made as high as 50 megohms. Many commercial circuits employ a total of only 10 megohms because resistors of this lower value are usually more stable than those of higher resistance.

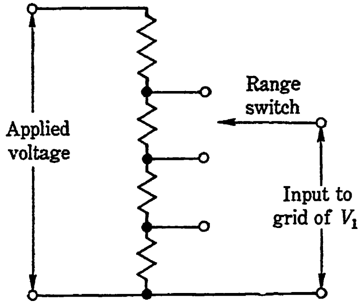


Fig. 16.4. A high-resistance voltage divider for extending the range of a d-c vacuum-tube voltmeter for applications in which the meter may be allowed to draw a small current.

A–C Input Filter. For many applications it is desirable to measure the direct component of a voltage in the presence of an alternating voltage. An alternating component small compared with the direct voltage causes no trouble because the circuit is linear and the meter reads only the average value of the current passing through it. A large alternating voltage, however, may cause the circuit to operate beyond the linear range to produce partial rectification and a shift in the meter reading. For this reason most vacuum-tube voltmeters employ a simple R–C filter at the input to reduce the alternating voltage reaching the grid of V_1. Figure 16.5 shows this filter with typical values of re-

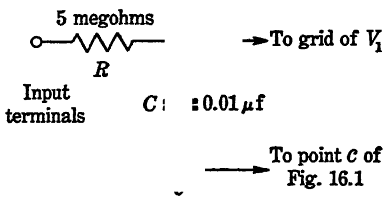


Fig. 16.5. An R–C filter to reduce the alternating component of voltage applied to the grid of the input tube.

sistance and capacitance. The analysis of Chap. 3 shows that this filter gives a smoothing factor of about 20 for the 60-cycle line frequency that is the most common source of trouble. Thus the filter reduces by a factor of 20 the 60-cycle hum picked up by the test leads when they are connected to a high-resistance circuit.

Of course filter capacitor C must be of excellent quality to en-

sure a very high leakage resistance. Ordinary paper dielectric capacitors are unstable in this respect, and a mica dielectric capacitor molded into a low-loss plastic with a voltage rating of about 1,000 volts makes a better choice.

16.2 An A–C Vacuum-tube Voltmeter

The average a-c meter requires a much larger current to operate the moving element than does a moving-coil d-c meter. Furthermore, such a meter remains accurate over a distinctly limited frequency range. Common power-type instruments lose accuracy above several hundred cycles per second, and the more sensitive types consisting of a d-c meter with a selenium or copper-oxide rectifier cover a frequency range extending only to 10 or 20 kilocycles. An a-c vacuum-tube voltmeter, on the other hand, draws very little current (although usually more than does a d-c vacuum-tube voltmeter), and it covers a very wide frequency range.

One type of a-c vacuum-tube voltmeter circuit consists of a stable amplifier followed by a diode rectifier which provides the current to operate a d-c meter. The input impedance of this arrangement consists essentially of the grid leak used for the first amplifier tube, in parallel with a shunting capacitance consisting of the tube input capacitance plus that of the wiring and input blocking condenser. The frequency response of the amplifier limits the useful frequency range, but modern wide-band design techniques now permit the construction of a single instrument covering a range from 10 cycles up to several megacycles with an accuracy of about 2 percent. With sufficient amplification the sensitivity may be very good; some instruments give full-scale deflection with an a-c input of only 3 millivolts.

Another less sensitive type of vacuum-tube voltmeter first rectifies the applied alternating voltage and then applies the rectified output to a d-c vacuum-tube voltmeter for measurement. This arrangement avoids the limitations of amplifiers and produces an instrument capable of handling frequencies from the low audio range up to several hundred megacycles or more. Figure 16.6 shows the type of *shunt* diode rectifier commonly used for this purpose. The circuit is similar to the simple half-wave *series* diode circuit of Fig. 3.1 in Chap. 3, except for having the load resistance connected in shunt with the diode instead of across the capacitor. This load resistance is 10 megohms or more to reduce

the current drawn from the source of alternating voltage. In operation, capacitor C recharges periodically at the peak of each alternating voltage cycle, and between times it slowly discharges

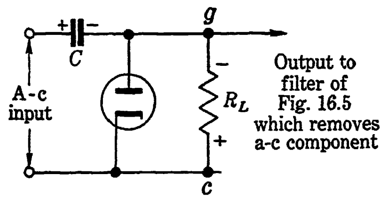


FIG. 16.6. A shunt diode rectifier that provides a d-c output proportional to the a-c input voltage. With a load resistance of many megohms this circuit draws very little current from the a-c scource.

through the load resistor in much the same fashion as the simple series diode circuit, except that the discharge is no longer exponential. However, with a sufficiently large value of C the high load resistance permits so little discharge during one cycle that the average voltage across C remains at practically the peak value of the input voltage, with the polarity shown. Since the voltage across the load resistor equals the sum of the input alternating voltage plus the substantially constant voltage across C, the wave form of the net voltage at g (with respect to point c) contains both an alternating and a direct component and looks like Fig. 16.7.

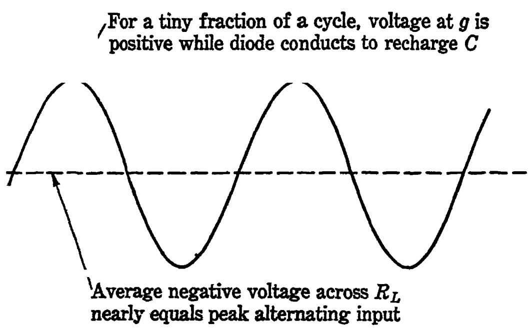


FIG. 16.7. Output-voltage wave forms for the shunt diode rectifier.

The rectifier output connects to an R–C filter to remove most of the a-c component before measurement by the d-c voltmeter circuit. Of course the d-c vacuum-tube voltmeter measures a value closely equal to the peak alternating voltage, but the con-

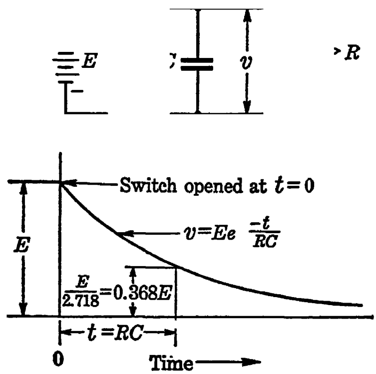
stant ratio between peak and effective value for a sine wave can be accounted for during calibration by the adjusting resistor R_m in the circuit of Fig. 16.1. Fortunately, the rectifier circuit remains linear over a range extending from about 1 volt up to several hundred so that a linear meter scale serves for the majority of readings. Below 1 volt the circuit becomes nonlinear, and a special scale must be used.

A serious disadvantage of this circuit is the fact that the readings are proportional to the peak amplitude of the alternating voltage. The usual calibration is adjusted to read 0.707 of peak voltage, which corresponds to the effective value for pure sine waves. Since the ratio between peak and effective value is different for other wave forms, the meter does not indicate the true effective value for complex waves. Furthermore, the meter responds only to the positive peaks of the wave which may be of different height than the negative peaks. In this case the meter will read a different value with the input connections reversed. This turnover error is a common fault of the majority of a-c vacuum-tube voltmeters.

APPENDIX A

THE R–C DISCHARGE

RECTIFIERS, timing circuits, and a number of other devices employ parallel resistance-capacitance circuits for the purpose of smoothing a voltage or producing a time delay. With these applications in mind we shall briefly study the behavior of the simple circuit shown by Fig. A.1.



$\tau = RC$

FIG. A.1. This shows the exponential discharge of a parallel resistance-capacitance circuit.

To analyze this circuit we start with three basic facts: (1) the voltage across the resistor is proportional to the current through it (Ohm's law); (2) at every moment the voltage v across the capacitor equals that across the resistor (Kirchhoff's law); and (3) the voltage across the capacitor is proportional to its charge. In equation form these relations are

$$v_r = iR \text{ (volts, amperes, ohms)} \quad \text{(A.1)}$$

$$v = v_r \text{ (volts)} \quad \text{(A.2)}$$

$$q = Cv \text{ (coulombs, farads, volts)} \quad \text{(A.3)}$$

The last of these is more useful after taking the time derivative and observing that dq/dt represents the capacitor charging current. This gives

$$i = \frac{dq}{dt} = C\frac{dv}{dt} \qquad \text{(charge)} \tag{A.4}$$

In this analysis we are interested in the *discharge* of C after opening the switch. Since a charging current is the opposite of a discharge, Eq. (A.4) becomes

$$i = -C\frac{dv}{dt} \qquad \text{(discharge)} \tag{A.5}$$

We shall now imagine that capacitor C has first been charged to voltage E with switch S closed. Then at zero time instant the switch opens to let C discharge through the resistor. From that moment on, the current i represents both the capacitor discharge current and the resistor current. Using this fact by substituting Eq. (A.5) into Eq. (A.1), we obtain

$$v_r = iR = -RC\frac{dv}{dt} \tag{A.6}$$

A substitution of this value of v_r into Eq. (A.2) gives

$$v = -RC\frac{dv}{dt} \tag{A.7}$$

This differential equation describes the circuit behavior, but it is more useful to integrate it and obtain an answer giving v as a function of the time. With Eq. (A.7) rearranged to separate the variables, we get

$$-\frac{dt}{RC} = \frac{dv}{v} \tag{A.8}$$

This is easily integrated to obtain

$$-\frac{t}{RC} = \ln v + K \tag{A.9}$$

where the symbol ln represents the natural logarithm and K is the constant of integration. To determine K we must insert the initial conditions that at zero time the voltage across C equals E. Thus

$$\begin{aligned} 0 &= \ln E + K \\ K &= -\ln E \end{aligned} \tag{A.10}$$

With the newly found value for K, Eq. (A.9) becomes

$$-\frac{t}{RC} = \ln v - \ln E \qquad \text{(A.11)}$$

$$= \ln \frac{v}{E}$$

From the definition of logarithm, if $a = \ln b$, then $b = e^a$, where e is the base of natural logarithms, 2.71828 . . . . Applying this to Eq. (A.11), we obtain

$$\frac{v}{E} = e^{-t/RC}$$

$$v = Ee^{-t/RC} \qquad \text{(A.12)}$$

This equation represents the curve shown in Fig. A.1. It is easy to insert zero for the time to check the initial value of E and to observe that as t becomes very large the exponential term approaches zero. The curve shows that the capacitor does not discharge at once, but instead the voltage gradually drops toward zero and, theoretically at least, never reaches the bottom.

To obtain some idea of the rate of voltage decay we must look to the exponent. Since an exponent must be dimensionless, the product RC represents time (ohms $\times$ farads = seconds). This factor is called the *time constant* of the circuit because it gives a handy indication of the speed with which the discharge takes place. Suppose we observe the voltage at a time RC seconds after opening the switch. Then

$$v = Ee^{-RC/RC} = Ee^{-1}$$

which says that at the end of RC seconds the voltage has dropped to about one-third of the initial value (actually $0.368E$). The diagram shows this to scale. At the end of five time constants the voltage drops to Ee^{-5} which is only $0.0068E$, or less than 1 percent of the initial value. Thus from a practical viewpoint the discharge is essentially complete after a time equal to $5RC$ seconds. A circuit with a capacitance of 2 microfarads and a resistance of 100,000 ohms has a time constant of $(2 \times 10^{-6})(10^5)$, or 0.2 second, and it discharges in about 1 second.

APPENDIX B

EFFECTIVE AND AVERAGE VALUES OF A SINE WAVE

THE MATHEMATICAL simplicity of a sine wave and the ease with which it can be represented by a rotating vector system has led to its adoption as the standard wave form for alternating-current power systems. Furthermore, any complex repeating wave can be represented as a Fourier series of sine waves; this makes the sine wave a fundamental building block of electrical circuit theory.

The basic equation of a sine wave is $y = Y_{\max} \sin \theta$, where $Y_{\max}$ is the peak or maximum value of the wave form. For a wave repeating in time with cycles of equal length, angle θ is proportional to the time interval. Hence

$$\theta = \omega t \tag{B.1}$$

with ω as the proportionality constant between time and angle. This makes the equation of a sinusoidal current wave have the form

$$i = I_{\max} \sin \omega t \tag{B.2}$$

Symbol f represents the frequency of the wave in number of cycles per second; each cycle produces an advance of 2π radians and at the end of one second the angle has progressed through a total of $2\pi f$ radians. Placing this information into Eq. (B.1) we find that $\omega = 2\pi f$ radians per second.

B.1 Effective Value of a Sine Wave

The effective value of a current or voltage wave is defined as equal to the direct current or voltage which would produce the same average heating of a resistor.

Let us imagine a repeating current wave of any shape $i(t)$ having a period or cycle length of T. This current passes through a resistor R and produces a power loss $p = R[i(t)]^2$. To obtain the

average power flow we integrate p over one cycle and divide by the base T.

$$P_{\text{average}} = \frac{1}{T}\int_0^T R[i(t)]^2\,dt \tag{B.3}$$

The effective value I of this current equals the steady current which would produce the same average heating. For a steady direct current the average and instantaneous power flow are the same.

$$P_{\text{average}} = RI^2 \tag{B.4}$$

Equating Eqs. (B.3) and (B.4) to find the relation between $i(t)$ and the effective value I, we obtain

$$RI^2 = \frac{R}{T}\int_0^T [i(t)]^2\,dt$$

Solved for I this gives

$$I = \sqrt{\frac{1}{T}\int_0^T [i(t)]^2\,dt} \tag{B.5}$$

In words this says that the effective value of a current wave equals the square root of the average square; for this reason effective value is also commonly called root-mean-square or rms value. A similar relation can be derived for voltage.

Applying Eq. (B.5) to Eq. (B.2) for a sine wave of current, and remembering that $T = 1/f$, we obtain

$$\begin{aligned} I &= \sqrt{f\int_0^{1/f} (I_{\max}^2 \sin \omega t)\,dt} \\ &= I_{\max}\sqrt{\frac{f}{2}\int_0^{1/f} (1 - \cos 2\omega t)\,dt} \\ &= I_{\max}\sqrt{\frac{f}{2}\left[1 - \frac{\sin 2\pi f t}{2\omega}\right]_0^{1/f}} \end{aligned} \tag{B.6}$$

Inserting the limits makes the sine term drop out and produces the result

$$I = I_{\max}/\sqrt{2} = 0.707 I_{\max}$$

Since all a-c meters read effective values it is necessary to multiply the reading by $\sqrt{2}$ to compute the peak value of a sine wave. The use of effective values has become standard and any

statement regarding the magnitude of a voltage or current refers to effective value unless specifically stated otherwise.

B.2 Average Value of a Half Sine-wave Loop

In the rectifier circuits of Chap. 3 and in many control circuits the wave shape consists of half sine waves. In such circuits only the average value or d-c component is of interest because the output passes through filters or smoothing circuits to remove the alternating components.

To find the average value of a half sine-wave loop we integrate to obtain the area under the loop and then divide by the base. Thus for a voltage wave of the form $e = E_{\max} \sin \omega t$

$$E_{\text{average}} = \frac{1}{\pi} \int_0^\pi E_{\max} \sin \omega t \, d\omega t$$

$$= -\left[\frac{E_{\max}}{\pi} \cos \omega t\right]_0^\pi = \frac{2}{\pi} E_m \tag{B.7}$$

Consequently the average value of a half sine-wave loop is $2/\pi$ or 0.636 times the maximum value.

INDEX